The Impact of Social Nudges on User-Generated Content for Social Network Platforms

Zhiyu Zeng[1], Hengchen Dai[2], Dennis J. Zhang[3], Heng Zhang[4], Renyu Zhang[5]
Zhiwei Xu, Zuo-Jun Max Shen[6]

[1] Tsinghua University, Beijing, China, ceng-zy13@tsinghua.org.cn
[2] University of California, Los Angeles, Los Angeles, CA, USA, hengchen.dai@anderson.ucla.edu
[3] Washington University in St. Louis, St. Louis, MO, USA, denniszhang@wustl.edu
[4] Arizona State University, Tempe, AZ, USA, hengzhang24@asu.edu
[5] The Chinese University of Hong Kong, Hong Kong, China, philipzhang@cuhk.edu.hk
[6] University of California, Berkeley, CA, USA, maxshen@berkeley.edu
*The order of the first author is based on contribution to this work, and the order of the next four authors is alphabetical.

Content-sharing social network platforms rely heavily on user-generated content to attract users and advertisers, but they have limited authority over content provision. We develop an intervention that leverages social interactions between users to stimulate content production. We study *social nudges* whereby users connected with a content provider on a platform encourage that provider to supply more content. We conducted a randomized field experiment (N$=993,676$) on a video-sharing social network platform, where treatment providers could receive messages from other users encouraging them to produce more but control providers could not. We find that social nudges not only immediately boosted video supply by 13.21% without changing video quality but also increased the number of nudges providers sent to others by 15.57%. Such production-boosting and diffusion effects, though declining over time, lasted beyond the day of receiving nudges and were amplified when nudge senders and recipients had stronger ties. We replicate these results in a second experiment. To estimate the overall production boost over the entire network and guide platforms to utilize social nudges, we combine the experimental data with a social network model that captures the diffusion and over-time effects of social nudges. We showcase the importance of considering the network effects when estimating the impact of social nudges and optimizing platform operations regarding social nudges. Our research highlights the value of leveraging co-user influence for platforms, and provides guidance for future research to incorporate the diffusion of an intervention into the estimation of its impacts within a social network.

Key words: Content Production, Platform Operations, Social Network, Field Experiment, Information-Based Intervention

---

1. Introduction

Online content-sharing social network platforms such as Facebook and TikTok, where users create and consume content, are playing an increasingly important role in society. As of January 2021, an estimated 4.2 billion people, 53.6% of the world's population, were using these platforms.[1]

[1] https://datareportal.com/reports/digital-2021-global-overview-report.

They have evolved into powerful marketing tools, reshaping the global economy. For example, advertising spending on these types of platforms is expected to reach US$230.30 billion in 2022.[2] User-generated content (UGC) on these platforms can exert considerable influence on consumer decision-making, affecting sales of products and services (see, e.g., Chen et al. 2011).

These platforms, by nature, rely heavily on UGC to engage and retain users and advertisers alike. However, since users who generate organic content ("content providers") are not paid workers and UGC is essentially a public good, platforms have limited control over how often and how much users produce content and at what quality level (Yang et al. 2010, Gallus 2017). Hence, the underprovision of UGC has been a challenge that interests both practitioners (Pew Research Center 2010) and academics (Burtch et al. 2018, Huang et al. 2019, Kuang et al. 2019). Understanding drivers of content production and devising effective operational levers to motivate content production are vital for content-sharing social network platforms—this is the focus of our research.

A prominent feature of these platforms is that users have intensive social interactions with each other. The platforms can leverage the connections between users to stimulate UGC supply, as well as to help solve other operational problems. We study a novel kind of intervention that utilizes existing connections between users, capitalizes on psychological principles about when people are motivated to exert effort, and contains no financial incentives. Specifically, we study *social nudges* implemented by a user's neighbors on a platform (i.e., platform users who are connected to this user) to explicitly encourage her to supply more content on the platform.[3] We propose that by taking the time to explicitly encourage the user to produce more, neighbors convey that they value the user and her existing work and at the same time communicate their interest in viewing more of the user's future content. This may make the user feel more competent and valued (Ryan and Deci 2000) and increase her confidence in her future work receiving continued appreciation, which further motivates content provision (Grant and Gino 2010, Bradler et al. 2016).

Prior psychological and management research suggests that recognition from managers, companies, or platforms (Ashraf et al. 2014a,b, Bradler et al. 2016, Banya 2017, Gallus 2017) can boost recipients' production and retention. However, scant research has causally examined the motivating power of pure *peer recognition* that is not accompanied by financial incentives; moreover, this limited work has presented mixed evidence for the effectiveness of peer recognition in boosting production (Restivo and van de Rijt 2014, Gallus et al. 2020). Also, prior research has been silent

[2] https://www.statista.com/outlook/dmo/digital-advertising/social-media-advertising/worldwide.

[3] The word *nudge* is a behavioral science concept for describing interventions that intend to change individuals' behaviors without altering financial incentives or imposing restrictions (Thaler and Sunstein 2009). Nudges are usually implemented by managers, marketers, and policy makers. We coin the term *social nudges* to refer to nonfinancial, nonrestrictive interventions that are intentionally implemented by neighbors within a social network to influence peers.

about how interactions on a platform and its underlying social network could reinforce the effects of an intervention on production. Taking a more holistic perspective, we implemented large-scale field experiments to not only estimate the direct effects of our intervention (social nudges) on recipients' content production but also assess how being exposed to the intervention facilitates the spread of the intervention, which further stimulates additional recipients' content production. We then incorporated empirical findings from these field experiments into a social network model to estimate the impact of our intervention on content production over the entire social network.

Specifically, we conducted two randomized field experiments on a large-scale video-sharing social network platform (hereafter "Platform O" to protect its identity). As on Facebook, each user on Platform O can play two roles: content provider and content viewer. Users can follow other users and be followed. In this setting, we refer to a user's followers and to the users whom the user herself follows as *neighbors*.

We study social nudges sent by one type of neighbor: a user's followers. For users involved in our experiments, their followers could send them a message to convey the interest in seeing their videos and nudge them to upload more videos. Users in our experiments were randomly assigned to either the treatment or the control condition. The only difference introduced by our experimental manipulation between the two conditions was whether users could actually receive social nudges: treatment users could receive social nudges sent by their neighbors but control users could not. Because the difference between the two groups of users lay in their roles as providers and our primary focus was content production, we hereafter refer to users involved in our experiments as *providers*. We conducted our main experiment—the focus of this paper—from September 12 to 14, 2018, and our second replication experiment from September 14 to 20, 2018.

Analyses about $993,676$ providers in our main experiment yield several important insights. To begin with, we present four main findings about the effects of social nudges on recipients' content production (**direct effects of social nudges on production**). First, receiving social nudges boosted the number of videos that treatment providers uploaded on the day they received the first nudges by 13.21%, without causing providers to alter their video quality. This in turn increased consumption of treatment providers' content by 10.42%. Second, receiving a social nudge yielded a larger immediate boost in production when a provider and the follower who sent the nudge had a two-way tie (i.e., the provider was also following the follower; 17.39%) than when they had a one-way tie (i.e., the provider was not following the follower; 9.37%), suggesting that stronger ties between users strengthen the effect of social nudges on production. Third, the effect of receiving social nudges on production declined over time but remained significant within three days of receiving social nudges (a relative increase of 13.21% on the day of receiving social nudges versus 5.29% and 2.54% on the first and second days afterwards, respectively). Fourth, leveraging data from another experiment on Platform O that studied nudges sent to providers *by the platform*, we find suggestive evidence that social nudges from peer users can more effectively boost production than platform-initiated nudges.

Next, we examine whether providers receiving social nudges became more likely to send nudges to users they follow, which, if holding true, could further boost production on the platform (**indirect effects of social nudges on production**). We present three key findings about nudge diffusion. First, treatment providers sent 15.57% more social nudges on the day of receiving social nudges relative to control providers. Second, receiving a social nudge had a stronger effect on providers' willingness to send social nudges when they got a nudge from a two-way tie (29.97%) vs. from a one-way tie (2.87%). Third, the diffusion effect of social nudges declined over time and was significant within two days of receiving social nudges (a 15.57% increase on the day of receiving social nudges versus a 7.87% increase on the following day).

The diffusion of social nudges by nudge recipients as well as the over-time effects of social nudges impose challenges for estimating the impact of social nudges on production and in turn optimizing platform operational strategies regarding social nudges in different scenarios. We refer to the stationary effect of social nudges on content production on the entire social network—where every user could receive and send social nudges—as the *global effect* of social nudges. To precisely estimate this effect, we propose an infinite-horizon stochastic social network model. We model the social network embedded on Platform O as a directed graph in which each user is a *node*, and each following relationship is an *edge*. Based on our empirical evidence, the actual number of nudges sent on an edge in a period depends on both (1) the baseline number of nudges that would be sent without the influence of nudge diffusion and (2) the number of nudges its origin has received (i.e., the diffusion of nudge). Each user's production boost in a period is determined by all the social nudges she has received.

We also incorporate the time-decaying effect of both direct and indirect effects of social nudges with estimated decaying factors. Leveraging such a social network model, we provide a framework to estimate the global effect of social nudges on production boost, and we show that simply comparing the number of videos uploaded by treatment vs. control providers right after they were sent social nudges during the field experiment severely underestimates the global effect of social nudges. Moreover, based on this model, we devise a variant of the Bonacich centrality for edges, and we further develop the social-nudge index of each edge that quantifies the total production boost attributed to this edge. Via simulation, we showcase that platforms can use a social-nudge index to optimize operational decisions such as optimal seeding and provider recommendation for new users, highlighting this model's potential to improve platform performance in various settings.

In summary, we study a low-cost, behaviorally informed intervention that is initiated by neighbors on online platforms and can be widely applied to content providers on a platform. Empirically, we document both its direct production-boosting effect and its diffusion by intervention recipients. Theoretically, we develop a model to incorporate its diffusion into a social network model, thus allowing for a precise estimate of its global effect on production over the entire platform, as well as optimization of its overall effectiveness. Methodologically, our work provides guidance to future researchers for more comprehensively estimating an intervention's causal effects on a social network. Practically, our proposed low-cost, psychology-based intervention is valuable to online content-sharing social network platforms for managing their UGC, and our model can be a useful tool for platforms to evaluate and optimize the strategy for increasing the global effect of an intervention on a social network.

The rest of the paper proceeds as follows. Section 2 reviews the relevant literature. Section 3 introduces our field setting, experimental design, and data. Sections 4 and 5 present the direct effects of social nudges on content production and the diffusion of nudges, respectively. Section 6 describes the social network model, counterfactual analyses, and two practical applications illustrating the operational implications of our model. In Section 7, we discuss practical implications of our research and directions for future research.

2. Literature Review

Our research builds primarily on four streams of literature: production, peer effects and social networks, information-based interventions, and platform operations.

Production. Our work is most closely connected to research that seeks to motivate content generation on online content-sharing platforms. The interventions examined in prior work include financial incentives (e.g. rewarding content providers with money; Cabral and Li 2015, Burtch et al. 2018, Kuang et al. 2019), social norms (e.g., informing content providers about what most of their peers do; Chen et al. 2010, Burtch et al. 2018), performance feedback (e.g., informing content providers about their performance; Huang et al. 2019), hierarchies (e.g., ranking content providers based on their contributions to a website; Goes et al. 2016), symbolic awards (e.g., giving content providers badges based on their recent activities; Ashraf et al. 2014a, Restivo and van de Rijt 2014, Gallus 2017), and a combination of these tools (Burtch et al. 2018, 2021).

Our contribution to this literature is threefold. First, we study a novel intervention (social nudges) that leverages individual-to-individual peer recognition, contains no material incentives, and is applicable to all content providers on a platform. Apparently, social nudges differ fundamentally from financial incentives, social norms, performance feedback, and hierarchies. And, although social nudges are related to symbolic awards in the sense that both convey recognition without monetary incentives, awards must be given to a select body of users who deserve them (e.g., users who recently contributed UGC, top-performing users) in order to maintain their prestige and meaning, and thus their scope is more limited than that of social nudges.

Second, the nascent literature that examines recognition-based interventions (Frey and Gallus 2017) has mostly studied recognition communicated by authoritative figures such as managers and organizations (Ashraf et al. 2014a, Gallus 2017). The scant work examining the causal effect of peer recognition without financial incentives (Restivo and van de Rijt 2014, Gallus et al. 2020) presents inconclusive evidence for whether peer recognition can increase users' contributions. Specifically, Restivo and van de Rijt (2014) conducted a field experiment among the top 10% of providers to Wikipedia. They found that peer recognition increased production only among the most productive 1% providers but did not affect other providers who were relatively less productive (those at the 91st to 99th percentile). If anything, the treatment *reduced* retention of providers at the 91st to 95th percentile. Such negative effect of peer recognition might occur because providers who were not the most prolific (e.g., those at the 91st to 99th percentile) did not see themselves as sufficiently qualified to receive the recognition, given that they had not received any recognition before and the recognition in the experiment came from experimenters who pretended to be peer users. In a field experiment among NASA's workforce, Gallus et al. (2020) found a null effect of peer recognition on individuals' contributions to a NASA crowdsourcing platform. Peer recognition may fail to motivate in this context because NASA employees did not perceive the recognized activity as part of their core work and thus did not view peer recognition as legitimate or meaningful. Thus, it remains an open question whether an intervention that conveys peer recognition can boost recipients' effort provision on a UGC social network platform. We speak to this open question by implementing large-scale field experiments to test the effectiveness of an intervention that conveys peer recognition.

Third, prior studies have focused on testing the effects of an intervention on targets' content production, but they have rarely focused on whether and how the intervention diffuses (i.e., how a user, upon receiving the intervention, spreads and applies it to influence other users). We take a critical first step in this direction by not only empirically examining the diffusion of social nudges but also incorporating the diffusion process into our social network model to more accurately estimate the impact of our intervention on content production over the entire social network.

Within the production literature, our research is also related to prior studies on how to lift productivity in service and manufacturing settings. These studies have focused on four types of interventions for increasing productivity: those that (1) are based on workers' economic considerations (Lazear 2000, Celhay et al. 2019), (2) offer workers training (De Grip and Sauermann 2012, Konings and Vanormelingen 2015) or introduce information technology (Tan and Netessine 2020),

(3) assign workers to various staffing or workload settings (Tan and Netessine 2014, Moon et al. 2018), and (4) capitalize on workers' psychological needs and tendencies (Kosfeld and Neckermann 2011, Roels and Su 2014, Song et al. 2018). These interventions are usually implemented by firms or managers. Extending this line of work, we develop and test a novel psychology-based intervention that does not originate from firms or managers but instead leverages peer recognition to motivate effort provision and production.

Peer Effects and Social Networks. Research about peer effects (Zhang et al. 2017, Bramoullé et al. 2020) often investigates how schoolmates (Sacerdote 2001, Whitmore 2005), coworkers (Mas and Moretti 2009, Tan and Netessine 2019), family members (Nicoletti et al. 2018), residential neighbors, and friends (Kuhn et al. 2011, Bapna and Umyarov 2015) affect someone's own behaviors, ranging from mundane consumption and product adoption to consequential outcomes about education, health, and career.

We extend this literature about peer effects in two ways. First, prior research usually estimates peer effects without distinguishing whether peers exert influence passively (e.g., peers' choices are observed by others who then feel pressure to choose accordingly) or actively (e.g., peers persuade others to make certain choices). We clearly assess the active impact of peers by examining a novel kind of interaction initiated by peers because of their intention to influence others (i.e., peers send nudges to others in the hope of boosting others' production). Second, whereas prior research has normally focused on the effects of peers' outcomes (or behaviors) on another person's outcomes (or behaviors) in the same domain, our work simultaneously examines how peers actively influence another person's production via sending a social nudge as well as how the nudged person subsequently "learns," adopts the same tactic, and spreads this form of active influence via sending nudges to more peers.

Besides peer effects, we also speak to the literature that optimizes operational objectives based on social network models such as identifying key users (Ballester et al. 2006), seeding (Zhou and Chen 2016, Candogan and Drakopoulos 2020, Gelper et al. 2021), pricing (Candogan et al. 2012, Papanastasiou and Savva 2017, Cohen and Harsha 2020), and advertising (Bimpikis et al. 2016). Drawing insights from this literature, we propose an infinite-horizon stochastic social network model to characterize user interactions in a social network that allows for the precise calculation and optimization of an intervention's global effect. Our work takes this literature one step further by leveraging causal estimates from field experiments to calibrate model parameters, leading to an end-to-end implementation of such an optimization strategy.

Information-Based Interventions. Our work adds to the emergent Operations Management literature that empirically tests the effectiveness of information-based interventions in solving operational problems. This literature has examined such interventions as offering customers more information about firms and the market (Buell and Norton 2011, Parker et al. 2016, Cui et al. 2019, Li et al. 2020, Mohan et al. 2020, Xu et al. 2021), and offering service providers more information about customers (Buell et al. 2017, Cui et al. 2020a, Dai et al. 2022). These interventions have been shown to increase customers' engagement with firms and perceived service value as well as to improve service speed and capacity. We contribute to this literature by designing a novel information-based intervention that originates from neighbors within a social network, then causally demonstrating its production-boosting effect and diffusion.

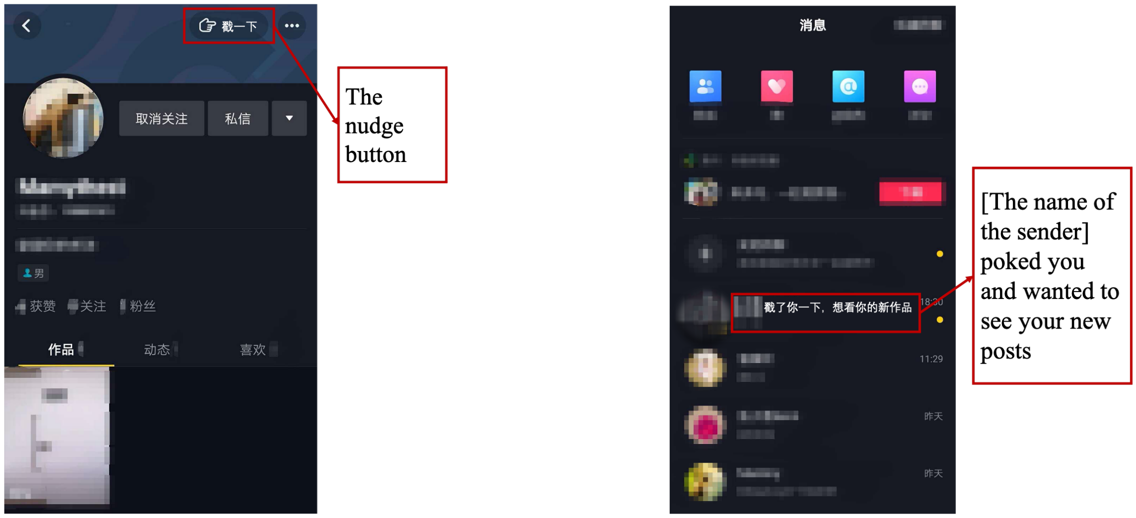
Platform Operations. Finally, our research extends the growing literature that addresses operations problems on online platforms. This literature has examined how to build effective systems for pricing (Cachon et al. 2017, Bai et al. 2019, Bimpikis et al. 2019, Zhang et al. 2020), recommendations (Banerjee et al. 2016, Mookerjee et al. 2017), staffing rules (Gurvich et al. 2019), and optimization of content production (Caro and Martínez-de Albéniz 2020); it has also studied how to estimate and leverage the spillover effects across platform users (Zhang et al. 2019, 2020), and how to ensure service quality (Cui et al. 2020b, Kabra et al. 2020). We contribute to this literature by empirically demonstrating that allowing platform users to send social nudges—a low-cost, easy-to-implement strategy—could lift content production and, in turn, total capacity and consumption on content-sharing platforms.

3. Field Setting, Experiment Design, and Data

3.1. Field Setting and Experimental Design

To empirically examine the impact of social nudges, we collaborated with Platform O, where each user can play two roles simultaneously—content provider and content viewer. Content providers (1) can upload videos for distribution on Platform O, (2) can decide when and what to post, and (3) do not get paid by Platform O for uploading videos. Content viewers can watch videos for free. Platform O, like most online content-sharing platforms, generates revenue primarily through online advertising (i.e., disseminating advertising videos to users).

Videos on Platform O are usually short, typically just a few seconds to a few minutes. Popular subjects include daily lives (e.g., views of a nearby park, work scenes, kids, pets), jokes or funny plots, performance (e.g., dancing, singing, making art), and know-how (e.g., cooking or makeup tips). Video content is usually displayed to users on one of three pages: (1) the page of videos uploaded by providers they follow, (2) that of popular videos recommended by Platform O, and (3) that of videos from providers who are geographically close to a given user.



(a) A Provider's Profile Page With the Nudge Button (b) A Social Nudge in the Message Center

Figure 1 How Social Nudges Are Sent by Neighbors and Displayed to Treatment Providers[4]

When watching a video, users can leave comments beneath the video and upvote it by clicking the *like* button. The only way for users to privately and directly communicate with each other on Platform O is through the private-message function. To establish closer relationships, users can follow others by clicking the "follow" button (available at the top of a video or on other users' profile page).

We conducted two randomized field experiments to causally test how social nudges from neighbors affected users' video production. Our first experiment lasted from 2 p.m. on September 11, 2018, to 5 p.m. on September 14, 2018. This is our main study. Our second field experiment, which replicates the first experiment, lasted from 5 p.m. on September 14, 2018, to the end of September 20, 2018. This experiment (see Online Appendix B for the data and results) targeted a smaller, nonoverlapping group of providers but lasted longer.

For providers involved in our experiments, their followers could send them a standard message to nudge them to upload new videos if they had not published videos for one or more days.[5] To do so, followers simply clicked a button on the provider's profile page that read, "Poke this provider" (*ChuoYiXia* in Chinese; see Figure 1 (a)). We refer to this behavior as "sending a social nudge."

Providers in our experiments were randomly assigned to either the treatment or the control condition. The *only* factor that we manipulated between the two conditions was whether providers

[4] To protect Platform O's identity, we digitally altered the app interface of a widely used video-sharing platform in China to obscure some nonessential details and reflect where the nudge button and social nudges are and what they look like on Platform O. Platform O has a similar app interface to Figure 1.

[5] Most providers could satisfy this requirement. For example, on the first day of the experiment, among all providers on Platform O who uploaded any videos in the past 30 days, 88% had not posted a video for one or more days.

were able to view social nudges sent to them. Specifically, treatment providers could see social nudges sent to them in their message center, along with other kinds of messages, whereas control providers could *not* see the social nudges in their message center. The standard social-nudge message to all providers said, "[Name of the sender] poked you and wants to see your new posts" (see Figure 1 (b)).[6] If treatment providers clicked on a social-nudge message, they would be directed to a list of all nudges that had ever been sent to them. On that page, newer nudges were displayed closer to the top. There, each social-nudge message read, "[Name of the sender] poked you [time when the nudge was sent] and wants to see your new posts." We designed these social nudges to be bare-bones simple, and standardized, so as to examine as cleanly as possible the basic effect of being nudged by a neighbor.

3.2. Data and Randomization Check

For the main analyses, our sample of providers (N $= 993{,}676$) included all treatment providers and control providers who satisfied two criteria: (1) at least one of their followers sent them a social nudge during our experiment, and (2) they had never received any social nudges before the experiment.[7] Treatment and control providers in our sample preserved the benefits of random assignment, because our random assignment of providers into the treatment condition vs. the control condition had no way of affecting whether and when their neighbors sent them the first social nudge during the experiment. To confirm the success of randomization among our sample of providers, we compared treatment providers (n $= 496{,}976$) and control providers (n $= 496{,}700$) in their gender, basic network characteristics, and preexperiment production statistics. As shown in Table 1, treatment and control providers in our sample had similar proportions of female providers, number of users who were following them ("number of followers") on the day prior to the experiment, and number of users they were following ("number of following") on the day prior to the experiment, as well as the number of videos they uploaded and the number of days when they uploaded any video during the week prior to the experiment. These results confirm that the treatment and control providers in our sample were comparable, suggesting that any difference between conditions after the experiment started should be attributed to our experimental manipulation—that is, whether providers could actually receive social nudges.

[6] In the message center, the most recent message appears at the top. Messages about social nudges were not given a higher priority over other types of messages. In general, messages disappear only when providers delete them.

[7] In the few months before our first experiment, social nudges were being tested and developed; as a result, some providers in our experiment received social nudges before the experiment. We removed those providers, per our second selection criterion, in order to estimate how social nudges change behavior when a platform starts to implement the social-nudge function. Our results are qualitatively unchanged if we remove the second criterion and include all providers whose followers sent them at least one social nudge during our experiment (see Online Appendix A.1).

Table 1 Randomization Check

		Treatment Providers (1)	Control Providers (2)	P-Value of Two-Sample Proportion Test or T-Test (3)
Statistics on the Day Prior to the Experiment	Proportion of Females	51.34%	51.38%	0.82
	Number of Followers	0.0622	0.0605	0.38
	Number of Following	0.8485	0.8480	0.81
Statistics During One Week Prior to the Experiment	Number of Uploaded Videos	0.3674	0.3693	0.33
	Number of Days with Videos Uploaded	0.5057	0.5078	0.30

Notes: All variables, other than whether a provider is a female, were standardized to have a unit standard deviation. To calculate the proportion of females, we excluded the $8,895$ providers ($\sim 0.9\%$) with missing gender information.

To protect Platform O's sensitive information[8], we standardized all continuous variables used in our analyses to have a unit standard deviation. To help readers better understand our empirical context, we report the scaled or standardized distributional information of relevant variables and network features in Online Appendix G. We also provide the code for our empirical and simulation analyses in a GitHub Repository[9].

4. Direct Effects of Social Nudges on Content Production

Our investigation began by examining the effects of receiving social nudges on the recipient's content production (i.e., the direct effects of social nudges on content production). The time unit we focused on was one day, which matches the granularity of our data offered by Platform O. Platform O cares about aggregate daily metrics (e.g., daily active providers, daily new videos), which breaks down to daily metrics at the individual level (e.g., on a given day, whether a user uploaded any video, how many videos she uploaded). In addition, 79% of providers in our sample had median intervals of video postings[10] at least one day, further confirming the appropriateness of using one day (rather than a smaller time window, such as one hour) as the time unit.

4.1. Direct Effects of Social Nudges on Content Production on the First Reception Day

We first tested whether social nudges had a positive effect on content production on the first day when a provider could be affected—that is, the day a provider was sent the first social nudge during the experiment; we refer to it as the providers' *first reception day*. Most (97%) providers in our

[8] The authors have a Non-Disclosure Agreement with Platform O.

[9] See `https://github.com/ZhiyuZeng-Public/the_impact_of_social_nudges_on_UGC`.

[10] For each provider, we calculated the interval (in days) between any two videos she successively uploaded (which equaled zero if two videos were uploaded on the same day) from January 1, 2018, to the day before the main experiment, and then we calculated her *median interval of video postings* across all pairs of successively uploaded videos.

Table 2 Direct Effects of Social Nudges on Content Production on the First Reception Day

	Main Treatment Effects			*Heterogeneous Treatment Effect*
Outcome Variable	Number of Videos Uploaded	Upload Incidence	Number of Videos Uploaded Conditional on Uploading Anything	Number of Videos Uploaded
	(1)	(2)	(3)	(4)
Treatment	0.0262^{****}	0.0094^{****}	-0.0168	0.0186^{****}
	(0.0020)	(0.0005)	(0.0181)	(0.0025)
Two-Way Tie				0.0700^{****}
				(0.0027)
Treatment * Two-Way Tie				0.0159^{***}
				(0.0041)
Relative Effect Size	13.21%	13.86%		
Observations	$993,676$	$993,676$	$71,883$	$993,676$

Notes: Continuous variables (Number of Videos Uploaded and Number of Video Uploaded Conditional on Uploading Anything) were standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on her first reception day. Columns (1), (2), and (4) include all providers in our sample. Column (3) includes the providers who uploaded at least one video on their first reception day. Robust standard errors are reported in parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

sample were sent only *one* social nudge on the first reception day, suggesting that the effects of our intervention on the first reception day were driven mostly by receiving one social nudge. Our unit of analysis was a provider on her first reception day; we analyzed $993,676$ observations, with each provider contributing one observation.

We used the following ordinary least squares (OLS) regression specification with robust standard errors to causally estimate the effects of social nudges on the first reception day:

$$Outcome\ Variable_i = \beta_0 + \beta_1 Treatment_i + \epsilon_i \tag{1}$$

where $Outcome\ Variable_i$ is detailed later and $Treatment_i$ is a binary variable indicating whether provider i was in the treatment (vs. control) condition.

For each provider i, we first examined the number of videos she uploaded on the first reception day ($Number\ of\ Videos\ Uploaded_i$). Column (1) of Table 2 reports the result of a regression that follows specification (1) to predict $Number\ of\ Videos\ Uploaded_i$. The positive and significant coefficient on treatment indicates that receiving social nudges immediately had a positive effect on the nudge recipient's production. Specifically, receiving social nudges increased the number of videos uploaded on the first reception day by 0.0262 standard deviations ($p < 0.0001$), a 13.21% increase relative to the average in the control condition.

Two underlying forces may drive this production-boosting effect: (1) providers became more willing to upload at least one video on the first reception day, and (2) providers who decided to upload at least one video on the first reception day uploaded more videos that day. To test the presence of the first force, for each provider i, we examined whether she uploaded at least one video on the first reception day ($Upload\ Incidence_i$). To test the presence of the second force, we examined the number of videos uploaded on the first reception day among providers who uploaded at least one video that day ($Number\ of\ Videos\ Uploaded\ Conditional\ on\ Uploading\ Anything_i$).

We used regression specification (1) to predict $Upload\ Incidence_i$ and $Number\ of\ Videos\ Uploaded\ Conditional\ on\ Uploading\ Anything_i$. Column (2) of Table 2 shows that receiving social nudges lifted the average probability of providers uploading any videos on the first reception day by 0.94 percentage points ($p < 0.0001$), a 13.86% increase relative to the average probability in the control condition. However, as shown in column (3) of Table 2, $Number\ of\ Videos\ Uploaded\ Conditional\ on\ Uploading\ Anything_i$ did not statistically significantly differ between conditions ($p = 0.3533$). Altogether, these results suggest that the boost in video supply on the first reception day was mainly driven by the first force—that is, providers became more willing to upload something after receiving social nudges.

Inspired by the social network literature (e.g., Jackson 2005), we next examined whether social nudges from closer peers could be more motivating. To answer this question, we tested whether the direct effects of social nudges on content production became stronger if a provider was also following the follower who sent her a nudge (in which case we refer to the relationship between the provider and the nudge sender as a two-way tie) than if the provider was not following that follower (in which case we refer to their relationship as a one-way tie). For each provider i on her first reception day, we identified the follower who sent the first social nudge to provider i, i.e., the *first social-nudge sender*. We constructed a binary variable, $Two\text{-}Way\ Tie_i$, which equals one if provider i was also following her first social-nudge sender and zero otherwise. We used the following regression specification with robust standard errors to predict $Number\ of\ Videos\ Uploaded_i$, where each observation was a provider on her first reception day:

$$\begin{aligned} Outcome\ \ Variable_i = & \beta_0 + \beta_1 Treatment_i + \beta_2 Two\text{-}Way\ \ Tie_i \\ & + \beta_3 Treatment_i * Two\text{-}Way\ \ Tie_i + \epsilon_i \end{aligned} \tag{2}$$

Column (4) of Table 2 shows that the coefficient on the interaction between $Treatment_i$ and $Two\text{-}Way\ Tie_i$ is significant and positive ($p < 0.001$). This suggests that, consistent with the social network literature (Jackson 2005), receiving social nudges increased a provider's content production to a greater extent when the provider and the follower who sent the nudge had a two-way tie than when they had a one-way tie. Specifically, receiving a social nudge from a follower with a one-way tie boosted the number of videos uploaded on the first reception day by 0.0186 standard deviations ($p < 0.0001$), whereas receiving a social nudge from a follower with a two-way tie boosted the number of videos uploaded by 0.0345 (i.e., 0.0186 + 0.0159) standard deviations ($p < 0.0001$). The relative effect size, compared to the average number of videos uploaded in the control condition, is 9.37% (one-way tie) and 17.39% (two-way tie), respectively.

4.2. Direct Effects of Social Nudges on Content Consumption and Content Quality

Beyond video production, how do social nudges affect overall video consumption and video quality? To evaluate the direct effects of social nudges on video consumption, we focused on the total number of views each provider engendered that could be attributed to videos they uploaded on the first reception day. Following Platform O's common practice, for each video uploaded on a provider's first reception day, we tracked the total number of views it received over the first week since its creation. Platform O normally uses the views each video accumulates during the first week after its creation to capture the short-term consumption it brings, because videos on Platform O are usually watched much more frequently during the first week and attract fewer views as time goes by. Then for each provider i, $\textit{Total Views}_i$ equals the total number of views within one week across all videos that provider i uploaded on the first reception day. If provider i did not upload videos on the first reception day, $\textit{Total Views}_i$ equals zero, which reflects the fact that no views were engendered by provider i as a result of her production effort on the first reception day. To address outliers, we winsorized $\textit{Total Views}_i$ at the 95th percentile of nonzero values.[11]

We used regression specification (1) to predict $\textit{Total Views}_i$. As shown in column (1) of Table 3, receiving social nudges increased the total views providers contributed to the platform as a result of their production effort on the first reception day by 0.0171 standard deviations, a 10.42% increase relative to the average in the control condition.[12]

To assess video quality, for every video uploaded by provider i on her first reception day, we collected four quality measures based on viewer engagement during the following week. Then for provider i, we calculated the average of each quality measurement across these videos: (1) the average percentage of times viewers watched a video until the end ($\textit{Complete View Rate}_i$), (2) the average percentage of viewers who gave likes to a video ($\textit{Like Rate}_i$), (3) the average percentage of viewers who commented on a video in the comments section beneath it ($\textit{Comment Rate}_i$), and (4) the average percentage of viewers who chose to follow provider i while watching a video ($\textit{Following Rate}_i$).

We used regression specification (1) to predict $\textit{Complete View Rate}_i$, $\textit{Like Rate}_i$, $\textit{Comment Rate}_i$ and $\textit{Following Rate}_i$. Columns (2), (4), and (5) of Table 3 Panel A indicate that social nudges did not significantly alter the complete view rate, comment rate, and following rate of videos uploaded on the first reception day (all p-values > 0.4). Column (3) suggests that videos uploaded

[11] Since the majority of providers produced no videos on the first reception day and consequently had a value of zero for $\textit{Total Views}_i$, the 95th percentile of the raw values of $\textit{Total Views}_i$ was small. Since we wanted to address extreme outliers caused by a small number of videos that went viral, we winsorized at the 95th percentile of nonzero values. That is, we replaced values of $\textit{Total Views}_i$ that were greater than the 95th percentile of nonzero values with the 95th percentile of nonzero values. The result is robust if we winsorize at the 99th percentile of nonzero values.

[12] The positive effect of social nudges on content consumption is robust if we use the total views a provider obtained on her first reception day (as opposed to within the first week of her first reception day) as the outcome variable.

Table 3 Effects of Social Nudges on Video Consumption and Quality

Panel A: Main Treatment Effects					
Outcome Variable	Total Views	Complete View Rate	Like Rate	Comment Rate	Following Rate
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0171****	0.0007	-0.0174*	-0.0068	0.0041
	(0.0020)	(0.0075)	(0.0075)	(0.0075)	(0.0075)
Observations	993,676	71,634	71,634	71,634	71,634
Relative Effect Size	10.42%		-1.48%		

Panel B: Investigating Why Treatment Providers Had Lower Like Rates Than Control Providers		
Outcome Variable	Historical Like Rate	Like Rate
	(1)	(2)
Treatment	-0.0522****	0.0081
	(0.0085)	(0.0062)
Historical Like Rate		0.5185****
		(0.0070)
Observations	69,825	69,594
Relative Effect Size	-3.48%	

Notes: All continuous variables were standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was the provider level. Column (1) in Panel A includes all providers in our sample. Columns (2) to (5) in Panel A include providers whose videos uploaded on their first reception day were watched at least once in the following week. Columns (1) and (2) in Panel B include providers whose videos uploaded on their first reception day were watched at least once in the following week and whose earlier videos were watched at least once between January 1, 2018, and the day prior to the experiment (September 11, 2018). Robust standard errors are reported in parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

by treatment providers on the first reception day were less likely to receive likes by 0.0174 standard deviations (1.48%) relative to videos uploaded by control providers ($p < 0.05$). To explore this difference in like rates, we further compared historical like rates between treatment and control providers who uploaded any videos on their first reception day. *Historical Like Rate*$_i$ equals the total number of likes provider i received from January 1, 2018, to the day prior to the experiment, divided by the total number of views provider i received during that same period.

Column (1) of Table 3 Panel B shows that among these providers who uploaded videos on the first reception day, treatment providers' historical like rates were significantly lower than control providers', by 0.0522 standard deviations (3.48%). This difference in historical like rates between treatment and control providers who uploaded videos on the first reception day could lead the like rates for videos uploaded on the first reception day to be lower in the treatment condition than in the control condition. In fact, when we predicted *Like Rate*$_i$ while controlling for *Historical Like Rate*$_i$, the coefficient on treatment was no longer significant (column (2) in Table 3 Panel B). Altogether, we find that social nudges did not *directly* cause providers to increase or decrease video quality.

4.3. Direct Effects of Social Nudges on Content Production Over Time

So far, we have shown that social nudges significantly lifted providers' willingness to upload videos on the first reception day, which in turn led them to contribute more views to the platform but did

Table 4 Over-Time Direct Effects of Social Nudges on Content Production

Outcome Variable	Number of Videos Uploaded			
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4
	(1)	(2)	(3)	(4)
Treatment	0.0262^{****}	0.0129^{****}	0.0065^{**}	0.0006
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Relative Effect Size	13.21%	5.29%	2.54%	
Observations	993,676	993,676	993,676	993,676

Notes: Number of Videos Uploaded was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on Day t relative to the first reception day, where $t = 1$ means the first reception day. Columns (1) to (4) include all providers in our sample. Robust standard errors are reported in parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

not change video quality. Next, we explored how the effect of receiving social nudges on content production changed over time. We compared the number of videos uploaded each day between treatment and control providers from the first reception day until the first day when the difference between conditions was not statistically significant. Specifically, for each day t starting from the first reception day (where t equals $1, 2, \cdots$ and $t = 1$ refers to the first reception day itself), we predicted the number of videos uploaded that day using regression specification (1).

Table 4 shows that the effect of receiving social nudges on content production was largest on the first reception day, decreased as time elapsed, but was positive and significant for a couple of days. Specifically, the number of videos uploaded was higher in the treatment condition than in the control condition by 13.21% on the first reception day (0.0262 standard deviations; $p < 0.0001$; column (1)), by 5.29% on the day after the first reception day (0.0129 standard deviations; $p < 0.0001$; column (2)), and by 2.54% on the second day after the first reception day (0.0065 standard deviations; $p < 0.0001$; column (3)). The effect of receiving social nudges on the nudge recipient's production was not significant on the third day after the first reception day ($p = 0.7644$; column (4)).

4.4. Additional Analyses About the Direct Effects of Social Nudges

This subsection is devoted to further discussions and analyses to supplement our main results.

Control Providers' Resentment. One potential alternative explanation for our observed difference in video production between treatment and control providers is that control providers somehow realized that they could not receive the social nudges sent by their followers, which made them resent the platform and thus reduce their production. Given that the private-message function is the only way for connected users to directly and privately communicate with each other on Platform O, this function is likely the only channel via which followers told control providers about social nudges they sent. Thus, we conducted two sets of additional analyses about the private-message function to address this alternative explanation (see Online Appendix C.1). First, we used the difference-in-differences (DiD) method to examine whether receiving private messages from followers who sent them social nudges during the experiment negatively affected control providers' content production. Second, we tested whether the treatment effect of social nudges on production differed between providers who received any private message from their first social-nudge sender during the experiment versus providers who did not. For both analyses, we find no evidence supporting the alternative explanation based on control providers' resentment.

Role of Likes and Comments. Since receiving social nudges could boost video production, nudge recipients might also receive more likes and comments due to the increased number of videos uploaded, which could in turn motivate nudge recipients to produce more. We tested how much the immediate increase in likes and comments due to the receipt of social nudges contributed to the effect of receiving social nudges on content production after the first reception day (see Online Appendix C.2). We find that increased number of likes and comments are neither the only reason nor the primary reason why the effect of receiving social nudges on content production lasted for days. Indeed, the magnitude of the production-boosting effect of social nudges after the first reception day was only decreased only by a slight to moderate amount when we controlled for the quantity of likes and comments providers obtained earlier in the experiment. This observation suggests that receiving social nudges per se is sufficient to boost video production beyond the first reception day, even without additional positive feedback from likes and comments.

Effects of Social Nudges Across Providers With Different Baseline Productivity. Restivo and van de Rijt (2014) found that a peer-recognition intervention motivated only the most productive 1% of content providers but not providers ranked at the 91st–99th percentile. We actually observe that receiving social nudges boosted production among the most productive 1% providers, the providers ranked at the 91st–99th percentile, and the providers ranked below the 91st percentile (see Online Appendix C.4). These results suggest that receiving social nudges is generally effective in motivating content provision across users with different levels of productivity.

Comparison With Platform-Initiated Nudges. To motivate content provision, a platform may also directly nudge its users. To explore whether social nudges from peers are more effective than nudges sent by the platform, we leveraged another randomized field experiment where content providers were randomly assigned to either receive or not receive nudges from Platform O (see Online Appendix C.5). Adopting similar empirical analyses as described in Sections 4.1 and 4.3, we find that social nudges boosted providers' production to a larger extent than platform-initiated nudges.

5. Indirect Effects of Social Nudges on Production via Nudge Diffusion

Going beyond social nudges' direct impact on content production, we next turn to the diffusion of social nudges. Inspired by the diffusion phenomenon in the social network literature (e.g., Zhou and Chen 2016), we focus on how receiving social nudges could affect the number of social nudges sent by the recipient to other providers they were following.

5.1. The Effects of Social Nudges on Nudge Diffusion on the First Reception Day

We began our investigation by testing how receiving social nudges facilitated nudge diffusion on the first reception day—the first day when a provider could be affected by social nudges during our experiment. Our unit of analysis was a provider on her first reception day, and we analyzed $993,676$ observations, with each provider contributing one observation. We examined the number of social nudges sent by each provider i to other providers on the first reception day (*Number of Social Nudges Sent*$_i$). Similar to how we addressed outliers earlier, we winsorized *Number of Social Nudges Sent*$_i$ at the 95th percentile of nonzero values. We used regression specification (1) to predict *Number of Social Nudges Sent*$_i$. Column (1) of Table 5 shows that on average receiving social nudges increased the number of social nudges providers sent to others on the first reception day by 0.0325 standard deviations (15.57%; $p < 0.0001$).

Next we tested whether social nudges from closer peers could more effectively facilitate nudge diffusion. Similar to how we analyzed the heterogeneous treatment effect for the direct production-boosting effect of social nudges (Section 4.1), here we examined the heterogeneous treatment effects for nudge diffusion based on whether a provider and the follower sending her a nudge had a two-way tie or a one-way tie. Specifically, we used regression specification (2) to predict *Number of Social Nudges Sent*$_i$.

Column (2) of Table 5 shows that the coefficient on the interaction between *Treatment*$_i$ and *Two-Way Tie*$_i$ is significant and positive ($p < 0.0001$), suggesting that receiving a social nudge motivated a provider to diffuse social nudges to a greater extent when the provider and the follower who sent the nudge had a two-way tie than when they had a one-way tie. Specifically, receiving a social nudge from a follower with a one-way tie boosted the number of social nudges a provider sent on the first reception day by 0.0060 standard deviations ($p < 0.05$), while receiving a social nudge from a follower with a two-way tie boosted the number of social nudges sent by 0.0625 (i.e., 0.0060+0.0565) standard deviations ($p < 0.0001$). The relative effect size, as compared to the average number of social nudges sent in the control condition, is 2.87% (one-way tie) and 29.97% (two-way tie), respectively. Combining these results with the findings in Section 4.1, we find that receiving social nudges both increased a provider's own content production to a greater extent and yielded a larger diffusion effect when the provider and the nudge sender were following each other

Table 5 Effect of Social Nudges on Nudge Diffusion on the First Reception Day

Outcome Variable	Number of Social Nudges Sent	
	(1)	(2)
Treatment	0.0325****	0.0060*
	(0.0020)	(0.0023)
Two-Way Tie		0.1304****
		(0.0028)
Treatment * Two-Way Tie		0.0565****
		(0.0041)
Relative Effect Size	15.57%	
Observations	993,676	993,676

Notes: Number of Social Nudges Sent was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on her first reception day. Columns (1) and (2) include all providers in our sample. Robust standard errors are reported in parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

than when only the nudge sender was following the provider, suggesting that social nudges from closer peers were more influential.

5.2. Effects of Social Nudges on Nudge Diffusion Over Time

Going beyond the first reception day, we next examined how receiving social nudges affected nudge diffusion over time. Similar to how we analyzed the direct effect of social nudges on content production over time, we compared the number of social nudges providers sent each day between treatment and control conditions from the first reception day on until the first day when the difference between conditions was not statistically significant. Specifically, for each day t starting from the first reception day (where t equals $1, 2, \cdots$ and t = 1 refers to the first reception day itself), we predicted the number of social nudges sent that day using regression specification (1).

Table 6 shows that the effect of receiving social nudges on the number of social nudges sent was largest on the first reception day and decreased as time elapsed. Specifically, the number of social nudges sent to others was higher in the treatment condition than in the control condition by 15.57% on the first reception day (0.0325 standard deviations; $p < 0.0001$; column (1)), and by 7.87% on the day after the first reception day (0.0139 standard deviations; $p < 0.0001$; column (2)). This effect of receiving social nudges on nudge diffusion was not significant on the second day after the first reception day ($p = 0.1686$; column (3)).

6. A Social Network Model

The reduced-form results reported in Sections 4 and 5 describe the transient and local impacts of social nudges. Platforms may be interested in evaluating the global effect of social nudges: the total impact of social nudges on production in the counterfactual scenario where *every* user on the platform can send and receive nudges. They may also be interested in optimizing various operational decisions regarding social nudges, such as seeding and recommending providers to new

Table 6 Effects of Social Nudges on Nudge Diffusion Over Time

Outcome Variable	Number of Social Nudges Sent		
	On Day 1 (First Reception Day)	On Day 2	On Day 3
	(1)	(2)	(3)
Treatment	0.0325^{****}	0.0139^{****}	0.0028
	(0.0020)	(0.0020)	(0.0020)
Relative Effect Size	15.57%	7.87%	
Observations	993,676	993,676	993,676

Notes: Number of Social Nudges Sent was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on Day t relative to the first reception day, where $t = 1$ means the first reception day. Columns (1) to (3) include all providers in our sample. Robust standard errors are reported in parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

users. However, the *over-time effects* and *diffusion* of social nudges, which we document in Sections 4 and 5, impose challenges for these tasks. To tackle these challenges, we propose a novel social network model to capture both the over-time effects and diffusion of social nudges. Applying this model allows us to quantify both the direct and the indirect effects of social nudges on content production over-time, and thus more accurately estimate the global effect of social nudges.

6.1. The Model and the Global Effect

We model Platform O as a social network, denoted as $G = (V, E)$, in which $V := \{1, 2, 3, \ldots, |V|\}$ is the set of nodes (i.e., users on Platform O who can be viewers and providers) and $E := \{1, 2, 3, \ldots, |E|\}$ is the set of directed edges (i.e., the "following" relationship on Platform O). We use i, j and e, ℓ to denote nodes and edges, respectively. Let e_o and e_d be the origin and destination, respectively, of edge $e \in E$, so viewer i following provider j is represented as $e = (i, j)$, $e_o = i$, and $e_d = j$. The dynamics of social nudges and their effects on providers' production are captured using a discrete-time stochastic model with an infinite time horizon. We use t to index the discrete time period (a single day in our empirical context, which is consistent with the business practice of Platform O), where $t = 1$ refers to the period when the social-nudge function first becomes available to all users on the platform. In Figure 2, we illustrate the structure of the social network model. If e_o sends e_d a nudge, the recipient, e_d, will not only (1) increase her production but also (2) send more nudges to other providers whom she is following, which could further boost other providers' production. We summarize the notations involved in the social network model in Table 7.

We first model the over-time direct effect of social nudges on production. Let $x_i(t)$ denote the boost of provider i's production in period t due to the social nudges she has received before and during period t. We use $y_e(t)$ to denote the number of nudges sent on edge e (from e_o to e_d) in period t. Let p_e denote the expected additional number of videos provider e_d would upload as a result of receiving one social nudge from viewer e_o on the day the nudge is received. Section 4 shows that, in our field experiment on Platform O, the direct effect of receiving social nudges on

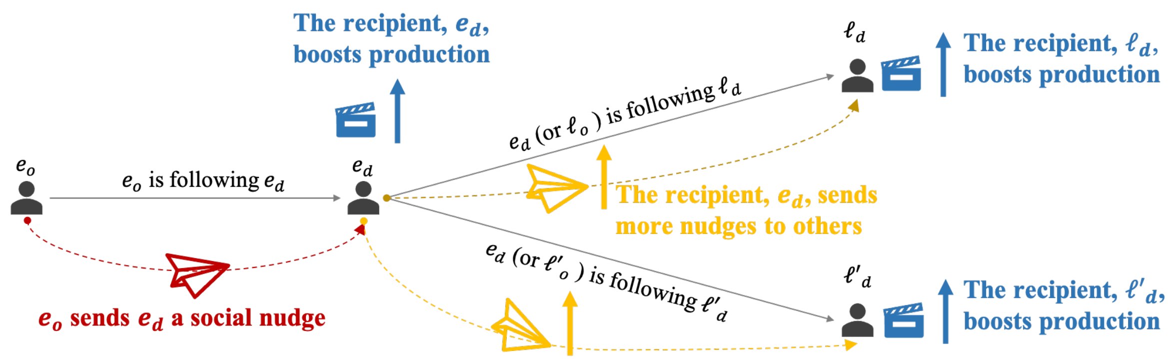


Figure 2 How Social Nudges Influence Users on a Network

Table 7 Notations Involved in the Social Network Model

Notations	Interpretations
$G=(V,E)$	The network, in which V is the set of nodes and E is the set of directed edges.
$x_i(t)$	The boost of node i's production in period t due to nudges node i has received before (including) period t.
$y_e(t)$	The number of nudges sent from e_o to e_d in period t.
p_e	The additional number of videos provider e_d would be expected to upload in period t as a result of receiving one social nudge from viewer e_o in period t.
μ_e	The number of nudges that e_o sends to e_d without being affected by the nudges that e_o has received.
$d_{\ell e}$	The expected increase in the number of nudges sent on edge e in period t due to one additional nudge e_o receives in period t from edge ℓ (i.e. $\ell_d=e_o$).
$\epsilon_i^x(t), \epsilon_i^y(t)$	The i.i.d. random noises with a zero mean and a bounded support.
α_p, α_d	The time-discounting factors corresponding to p_e and $d_{\ell e}$, respectively.

production gradually wears off over time. Thus, we capture the dynamic of production increment by the following dynamic equation:

$$x_i(t)=\sum_{1\le s\le t}\alpha_p^{t-s}\sum_{e\in E:e_d=i}p_e y_e(s)+\epsilon_i^x(t),\ \forall i\in V \tag{3}$$

where $\alpha_p\in(0,1)$ denotes the time-discounting factor of social nudges' direct production-boosting effect. We denote the random noise of production boost for provider $i\in V$ in period t as $\epsilon_i^x(t)$, independent across different providers and periods with zero means.

We next model the diffusion of social nudges. Motivated by the empirical results in Section 5, we assume that the number of social nudges sent on an edge e in period t is driven by two additive factors. First, we let μ_e denote the expected number of nudges sent on edge e that are not affected by the number of nudges e_o herself has received. We refer to μ_e as the expected number of *organic nudges* and denote $\boldsymbol{\mu}:=(\mu_e:e\in E)$. Second, the diffusion effect described in Section 5 suggests that when a provider receives a nudge, she tends to send more nudges to other providers she follows. We refer to these social nudges engendered through the diffusion process as *diffused nudges*. Combined, the dynamic of social nudges on the network G is captured by

$$y_e(t) = \mu_e + \sum_{1\le s\le t} \alpha_d^{t-s} \sum_{\ell\in E:\ell_d=e_o} d_{\ell e} y_\ell(s) + \epsilon_e^y(t),\ \forall e\in E. \tag{4}$$

Here the second term in Equation (4) embodies the diffusion effect. In particular, $d_{\ell e}$ captures the intensity of social-nudge diffusion, i.e., the expected increase in the number of nudges sent on edge e in a given period due to one additional nudge e_o receives in the same period on edge ℓ directing to e_o (i.e. $\ell_d = e_o$). Similar to α_p, $\alpha_d \in (0,1)$ denotes the time-discounting factor of nudge diffusion, which captures the extent to which the diffusion effect resulted from a single nudge decays over time, as discussed in Section 5. We denote the random noise of social nudges sent on edge e in period t as $\epsilon_e^y(t)$, independently distributed across different edges and periods with zero means.

Equations (3) and (4), built on the well-established models to study social interactions in the literature (e.g., Ballester et al. 2006, Candogan et al. 2012, Zhou and Chen 2016) and the key empirical observations from our experimental data, are the backbones of our social network model and together capture the over-time effects and diffusion of social nudges. As we will show below in Section 6.2 and Online Appendix E.4, both the estimation of the model parameters (Table 8) and that of different terms (the direct and indirect effects) in the global effect of social nudges (Table 9) are fairly consistent with respect to data from different experiments on Platform O. Such consistency provides further evidence that our model could reasonably capture the interactions observed in our network data.

To quantify the global effect of social nudges, we characterize the long-run steady state of the system defined by Equations (3) and (4). Theorem 1, whose proof is in Online Appendix D.2, shows that the expected production and nudge quantities converge to a well-defined limit. We define $d_{\ell e} = 0$ if $\ell_d \neq e_o$, and the matrix $\mathbf{D} := (d_{\ell e} : (\ell, e) \in E^2)$. The matrix $\mathbf{D}$ with nonnegative entries therefore captures the first-order diffusion on all edge pairs of the social network. We further define $\eta_e := p_e/(1-\alpha_p)$ and $\boldsymbol{\eta} := (\eta_e : e \in V)$. We use $\mathbf{I}$ to denote the identity matrix of appropriate dimension. The total production increment in period t is $x(t) := \sum_{i\in V} x_i(t)$. Define a matrix series:

$$\mathbf{M}(k) := \mathbf{I} + \sum_{i=1}^{k} \frac{1}{(1-\alpha_d)^i} \mathbf{D}^i,\ \text{for } k\in\mathbb{Z}_+.$$

A key condition we need here is the convergence of $\mathbf{M}(k)$ to a finite-valued matrix, as $k\to\infty$. In this case, we say that $(\alpha_d, \mathbf{D})$ satisfies **Condition** $\mathcal{C}$. Note that, since $\mathbf{D}$ is nonnegative, $\mathbf{M}(k)$ is componentwise increasing in k, so $\lim_{k\to+\infty} \mathbf{M}(k)$ is well-defined if and only if $\mathbf{M}(k)$ is componentwise bounded from above. Also, note that Condition $\mathcal{C}$ holds if the ℓ_∞ matrix norm of $1/(1-\alpha_d)\mathbf{D}$ is strictly below one (Horn and Johnson 2012). Indeed, for the real social network of Platform O, we verify that $||1/(1-\alpha_d)\mathbf{D}||_\infty < 1$, which implies Condition $\mathcal{C}$ holds (see Online Appendix D.1 for details). Inspired by the classical Bonacich centrality measure defined for nodes in the network economics literature (e.g., Ballester et al. 2006), we define the following *Bonacich centrality for edges.*

Definition 1 *Given the social network G and the associated diffusion matrix $\mathbf{D}$, we define the Bonacich centrality for edges (BCE) measure on E with respect to vector $\boldsymbol{v}$ as*

$$\mathcal{BE}(\mathbf{D},\boldsymbol{v}) := \left(\mathbf{I} - \frac{1}{1-\alpha_d}\mathbf{D}\right)^{-1}\boldsymbol{v} \tag{5}$$

where $\boldsymbol{v}$ is real-valued with compatible dimension, provided that $(\alpha_d, \mathbf{D})$ satisfies Condition $\mathcal{C}$.

We remark that Condition $\mathcal{C}$ guarantees that $\mathbf{I} - (1/(1-\alpha_d))\mathbf{D}$ is invertible,[13] so $\mathcal{BE}(\mathbf{D},\boldsymbol{v})$ is well-defined for any $\boldsymbol{v}$. The following theorem shows that the global effect of social nudges in the long-run steady state can be characterized by the BCE measure.

Theorem 1 *If $(\alpha_d, \mathbf{D})$ satisfies Condition $\mathcal{C}$, it then follows that $\lim_{t\to\infty} \mathbb{E}[x(t)] = x^*$ and $\lim_{t\to\infty} \mathbb{E}[\boldsymbol{y}(t)] = \boldsymbol{y}^*$, where x^* and $\boldsymbol{y}^*$ satisfy $x^* = \boldsymbol{\eta}^\top \boldsymbol{y}^*$ and*

$$\boldsymbol{y}^* = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu}). \tag{6}$$

In brief, Theorem 1 takes into account the over-time effects and the diffusion of social nudges. Importantly, for any $e \in E$, the BCE measure $\mathcal{BE}_e(\mathbf{D},\boldsymbol{\mu})$ quantifies the total expected number of nudges user e_o sends to e_d, including both the organic nudges and the diffused nudges. The factors $1/(1-\alpha_d)$ in Equation (5) and $1/(1-\alpha_p)$ in the definition of $\boldsymbol{\eta}$ materialize the diffusion and production-boosting effects, respectively, that accumulate over time. As we will show in Section 6.2, under Condition $\mathcal{C}$, the BCE measure bears a natural expansion with a clear economic interpretation that $\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ can be decomposed according to the radius of nudge diffusion.

6.2. Approximation and Estimation of the Global Effect

By Equation (5), an exact evaluation of the global effect of social nudges on providers' production involves inverting the $|E|^2-$dimensional matrix $\mathbf{I} - (1/(1-\alpha_d))\mathbf{D}$. For Platform O, the dimension of $\mathbf{I} - (1/(1-\alpha_d))\mathbf{D}$ is roughly at the magnitude of 10^{32}, so its inverse is computationally infeasible to obtain. Therefore, we resort to an approximation scheme to quantify the steady-state (daily) number of social nudges between viewers and providers (i.e., $\boldsymbol{y}^*$), and the (daily) production boost from these nudges (i.e., x^*).

[13] See Lemma 2 in Online Appendix D.1 for a formal proof.

Toward this goal, we note, by Lemma 2 (in Online Appendix D.1), that if $(\alpha_d, \mathbf{D})$ satisfies Condition $\mathcal{C}$, the inverse of $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D}$ is given by $\mathbf{I}+\sum_{i=1}^{\infty}(1/(1-\alpha_d)^i)\cdot\mathbf{D}^i$ (Equation (14) in Online Appendix D.1). Motivated by this formula, we define a sequence of (approximate) BCE measures, indexed by $k \in \mathbb{Z}_+$, as

$$\widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{v}, k) := \mathbf{M}(k)\cdot \boldsymbol{v} = \left(\mathbf{I} + \sum_{i=1}^{k} \frac{1}{(1-\alpha_d)^i}\mathbf{D}^i\right)\boldsymbol{v}. \tag{7}$$

Thus, we can develop approximates of the steady-state social-nudge vectors, $\tilde{\boldsymbol{y}}(k)$, and total production boost from nudges, $\tilde{x}(k)$:

$$\tilde{\boldsymbol{y}}(k) := \widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}, k), \text{ and } \tilde{x}(k) := \boldsymbol{\eta}^{\top}\tilde{\boldsymbol{y}}(k). \tag{8}$$

The following result, which is a corollary of Lemma 2 and Theorem 1, validates using $\tilde{\boldsymbol{y}}(k)$ and $\tilde{x}(k)$ to approximate $\tilde{\boldsymbol{y}}^*$ and $\tilde{x}^*$, respectively.

Corollary 1 *Assume that $(\alpha_d, \boldsymbol{D})$ satisfies Condition $\mathcal{C}$. We have (a) $\lim_{k\uparrow+\infty}\tilde{\boldsymbol{y}}(k) = \boldsymbol{y}^*$ and $\lim_{k\uparrow+\infty}\tilde{x}(k) = x^*$; (b) $\tilde{y}_e(k)$ is increasing in k for any $e \in E$, and so is $\tilde{x}(k)$ increasing in k. Therefore, for each $k \in \mathbb{Z}_+$, $\tilde{y}_e(k) \le y_e^*$ for all $e \in E$), and $\tilde{x}(k) \le x^*$.*

Economically, the approximate BCE, $\widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}, k)$, is the expected total number of nudges sent on each edge in E if the diffusion radius is at most k. Because the diffusion matrix $\mathbf{D}$ has an extremely high dimension, we introduce two important approximations to make the estimation of the global effect of social nudges computationally tractable. First, we adopt the approximation scheme (8) with $k=1$, thus ignoring the effect of nudge diffusion beyond radius one. As we will show below, such approximation will only incur a relative error of less than 1% for the global effect of social nudges on Platform O. Second, we adopt another layer of approximation by downsampling a subset of providers from V (denoted as $\tilde{V}$). We estimate the total production boost of the providers in $\tilde{V}$ brought by the social nudges they receive, denoted as $\hat{w}_0$, as well as the total production boost caused by the social nudges the providers in $\tilde{V}$ send out *as a result of the social nudges they receive* (i.e., the diffusion of nudges), denoted as $\hat{w}_1$. Hence, $\hat{w}_0$ captures the direct effect of social nudges and $\hat{w}_1$ captures the indirect effect in the steady state per period. Both $\hat{w}_0$ and $\hat{w}_1$ take into account the over-time effects of social nudges. Scaling these estimates by a factor of $\frac{|V|}{|\tilde{V}|}$ would therefore yield unbiased estimates of the true direct and indirect global effects. Therefore, we devise $\frac{|V|}{|\tilde{V}|}(\hat{w}_0+\hat{w}_1)$ as an unbiased estimate for $\tilde{x}(1)$.[14] We summarize the detailed estimation procedure as Algorithm 1 in Online Appendix D.3.

[14] See Proposition 1 for a formal proof.

Based on Algorithm 1, quantifying the global effect for Platform O involves estimating the following four sets of parameters: (1) the expected number of organic social nudges for each edge, i.e., μ_e for $e \in E$; (2) the effect of receiving one social nudge on boosting the nudge recipient's production, i.e., p_e for $e \in E$; (3) the intensity of social-nudge diffusion, i.e., $d_{e\ell}$ for $e, \ell \in E$ and $e_d = \ell_o$; and (4) the time-discounting factors, i.e., α_p and α_d. Our estimation of μ_e is based on observational data, whereas that of p_e, $d_{e\ell}$, α_p, and α_d relies on experimental data. The estimation results of the model parameters based on data from different experiments are provided in Table 8. We relegate the estimation details to Online Appendix E.

Table 8 Estimation of Parameters in the Social Network Model

Parameter	Estimation Results Using Data From the Main Experiment (1)	Estimation Results Using Data From the Replication Experiment (2)
p_e	0.05492	0.05156
α_p	0.6345	0.6945
$d_{e\ell}$	0.0008436	0.0009200
α_d	0.3750	0.3378

Notes: To protect Platform O's sensitive information, we are not permitted to disclose the raw estimates of p_e and $d_{e\ell}$. The values of p_e and $d_{e\ell}$ reported here equal the raw estimates of p_e and $d_{e\ell}$ multiplied by a fixed constant. We report α_p and α_d using the raw estimates.

Before presenting the estimate for the global effect of social nudges on production using Algorithm 1, we first describe a naïve benchmark that directly uses data from our experiment to calculate the difference in the number of videos uploaded by treatment vs. control providers on the first day when they are sent a social nudge. Then, we scale this difference to the entire population on the platform by the average number of providers who are sent social nudges on the platform per day, which can be estimated by (1) the number of providers in the analysis sample of our experiment who received social nudges on a day, divided by (2) the ratio of the number of providers targeted by the experiment to the total number of providers on the platform.

Following the above naïve approach and using data from our main experiment, we first estimate that the total boost of video uploads caused by social nudges among $1,000,000$ providers is 48.65 per day. Then following Algorithm 1, we approximate the total production boost of social nudges on the entire network on a given day in the steady state by downsampling a subset of providers $\tilde{V}$ where $|\tilde{V}| = 1,000,000$. To protect sensitive data, we only report the boost on $\tilde{V}$ without rescaling it back to the entire platform (i.e., $\hat{w}_0 + \hat{w}_1$). The estimation results using data from the main experiment are presented in Table 9 column (1). For those $1,000,000$ randomly sampled providers in $\tilde{V}$, the accumulated direct production boost is $\hat{w}_0 = 130.08$ videos per day, and the accumulated indirect production boost from social-nudge diffusion is $\hat{w}_1 = 10.59$ videos per day, yielding a total production boost of $\hat{w}_0 + \hat{w}_1 = 140.67$ videos per day. Therefore, our results suggest that the indirect production boost from nudge diffusion accounts for at least 8.14% of the direct effect (i.e., 10.59/130.08).

In addition, we remark that the estimation results discussed above suggest that using $\tilde{x}(1)$ is a reasonable approximation of x^*. Specifically, since the (first-order) indirect effect from nudge diffusion is about 8.14% of the direct effect, the production boost from second- and higher-order diffusion accounts for only about 0.72% (i.e., $\frac{0.0814^2}{1-0.0814}$) of the direct effect. Thus, ignoring the diffusion with radius two or beyond will introduce only fairly small additional errors.

Table 9 Estimation of the Global Effect of Social Nudges

	The Naïve Approach Using Data From the Main Experiment	The Network-Modeling Approach Using Data From the Main Experiment		The Network-Modeling Approach Using Data From the Replication Experiment	
	(1)	(2)		(3)	
Direct Effect	48.65	130.08	*One Day:* 47.55 *Beyond One Day:* 82.53	146.06	*One Day:* 44.63 *Beyond One Day:* 101.44
Indirect Effect		10.59		12.24	
Global Effect		140.67		166.30	
Ratio of Indirect Effect to Direct Effect		8.14%		8.38%	

Notes: When reporting the direct effect estimated by the network-modeling approach, we present the estimated overall direct effect over time (e.g., 130.08 for the first experiment), and we separately show the estimated direct effect on the day of receiving nudges (e.g., 47.55) and the estimated direct effect beyond that day (e.g., 82.53).

It is clear that our social network model could help address the substantial underestimate of the naïve approach to predict the social nudges' total production boost. The more precise estimation of social nudges' global effect over the entire user population using our social network model (140.67 per day for $1,000,000$ providers) is 2.89 times as large as the naïve estimate (48.65 per day for $1,000,000$ providers). Such a huge gap comes from two factors: (1) the social network model incorporates the over-time accumulation of the direct boosting effect of social nudges on recipients' production, which yields a 167% (i.e., (130.08 − 48.65)/48.65) increase compared to the naïve estimation; and (2) the model also captures the diffusion of nudges, which accounts for another 22% (i.e., 10.59/48.65) increase. We obtain similar results based on data from the replication experiment, as shown in Table 9 column (2). This robustness check, along with another one based on a different random sample of $\tilde{V}$ (see Online Appendix E.4), confirms the robustness of our estimation and validates the accuracy of our model in quantifying the global effect of social nudges on production boost on Platform O. Above all, our social network model provides a framework to causally quantify the global effect of our intervention (including its direct and indirect effects), which will be underestimated by the naïve estimation method.

6.3. Operational Implications

In this section, we demonstrate the operational implications of our social network model with two important practical applications: (1) seeding and targeting for the social nudge function, and (2) recommendation of content providers to new users. To this end, we first leverage the BCE measure to construct a *social-nudge index* (SNI) that assigns a metric to each (existing or new) edge that quantifies its value in production boost through social nudges.

For each edge $e \in E$, we define its SNI as the expected per-period total production boost on the entire network that can be attributed, either directly or indirectly through diffusion, to the *organic nudges* sent by e_o to e_d. Denote $\boldsymbol{\mu}_e \in \mathbb{R}^{|E|}$ as a vector with all entries equal to 0, except for that of edge $e \in E$ being μ_e. Define the SNI of edge $e \in E$ as:

$$\nu_e := \boldsymbol{\eta}^T \cdot \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}_e), \text{ provided that } (\alpha_d, \mathbf{D}) \text{ satisfies Condition } \mathcal{C} \tag{9}$$

where $\mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}_e)$ is given in Definition 1. As discussed above, exactly computing $\mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}_e)$ is computationally infeasible for a large-scale social network such as Platform O. Instead, we can bound ν_e from below, leveraging the approximate BCE as follows:

$$\tilde{\nu}_e(k) := \boldsymbol{\eta}^T \cdot \widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}_e, k), \text{ provided that } (\alpha_d, \mathbf{D}) \text{ satisfies Condition } \mathcal{C} \tag{10}$$

where $\widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}_e, k)$ is given by Equation (7). Similar to evaluating the global effect of social nudges, we focus on the case $k = 1$ in the computational simulation to balance accuracy and tractability. Therefore, of particular importance is the approximate SNI with diffusion radius $k = 1$ (so diffusion of order two or higher is ignored):

$$\tilde{\nu}_e(1) = \boldsymbol{\eta}^T \cdot \widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}_e, 1) = \frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell:\ell_o = e_d} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}, \text{ for } e \in E. \tag{11}$$

The approximate SNI (i.e., Equation (11)) offers insights on the property of a high-value edge: it either generates a high volume of organic nudges (the first term) or promotes a high volume of diffusion (the second term). For a wide range of practical applications, the key is to target the edges on a social network whose organic nudges boost provider production over the entire platform the most. With our social network model, this problem is equivalent to selecting the edges in E with the highest social-nudge indices. In the case in which computing the (exact) SNIs is intractable, we can further reduce this problem to a simpler one of finding the edges $e \in E$ with the largest $\tilde{\nu}_e(1)$'s as a reasonable approximation. Next, we briefly illustrate how (approximate) SNIs can be used to address the seeding problem and the content-provider recommendation problem for content-sharing social network platforms. The details are relegated to Online Appendix F.

Optimal Seeding. To boost content production, a content-sharing platform may use operational levers to prompt users to send social nudges. For example, the platform can use push notifications or private messages that encourage viewers to send out social nudges to specific providers. Sensibly, any type of operational lever would require user attention while users only have limited attention and patience (Dukas 2004). Therefore, the platform must carefully control the intensity of such interventions to avoid disturbing or upsetting its users.

Considering the limited number of levers that the platform could use at once without causing annoyance, the usage of one lever means forgoing the opportunity of implementing another lever. In this sense, when seeking to get more viewers to send out social nudges, the platform is faced with a capacity constraint, has to decide on which edge to exert influence via a given lever, and has to select a set of n edges $K \subset E$ to target. We denote that for each $e \in K$, the average number of social nudges sent on this edge per day will increase by a relative effect of δ_μ after e_o receives the motivation from the platform (i.e., from μ_e to $\mu_e(1+\delta_\mu)$). The platform could control the strength of its encouragement for users to send more social nudges by adopting the appropriate lever. In our model, this is captured by the platform being able to change the parameter δ_μ according to its need. For example, besides targeting push notifications or private messages to selected viewers, the platform can modify the app user interface of some viewers to highlight the social-nudge function for certain providers they are following. Based on our conversation with Platform O, the latter approach is likely to have a greater impact on users' behavior but requires much greater resources to set up, compared to the former one.

Next, we explore how the platform should optimize the global effect of social nudges, and estimate the extent to which the optimal strategy outperforms a random dissemination strategy in increasing the global effect of social nudges.

The global effect of social nudges with respect to the selected edges, K, is $\boldsymbol{\eta}^T \mathcal{B}\mathcal{E}(\mathbf{D}, \boldsymbol{\mu}_K)\delta_\mu$, where $\boldsymbol{\mu}_K \in \mathbb{R}^{|E|}$ represents a vector with an entry of edge $e \in K$ (resp. $e \notin K$) equal to μ_e (resp. 0). Such production boost can be reasonably approximated by $\delta_\mu \cdot \sum_{e\in K} \tilde{\nu}_e(1)$. Thus, it is (approximately) "optimal" to select n edges in E with the highest (approximate) SNIs, i.e., the n edges with the largest $\tilde{\nu}_e(1)$'s. As a benchmark, the platform may adopt the simple, straightforward strategy of randomly targeting a subset of edges $K \subset E$ ($|K| = n$) and encouraging the users to nudge more on these edges, i.e., the random strategy. By simulation, we calculate the relative improvement of the "optimal" strategy over the random strategy in the total production boost of social nudges. We find that the "optimal" strategy substantially outperforms the random strategy regardless of the effectiveness of the platform's encouragement for users to send additional nudges, δ_μ, especially when the size of selected target providers n is small. See Online Appendix F.1 for details.

Content Provider Recommendation for New Users. An important strategy for a platform to engage and retain newly registered users is to recommend to them some providers who they can follow and potentially nudge afterwards. Considering users' limited attention, the platform needs to decide the ranking of the provider list, after which it sequentially recommends the listed content providers to new users. After receiving the recommended list of providers, a new user may follow some or all of them. These new following links will in turn enable the new user to send social nudges to these providers and boost their content production. The platform seeks to maximize the total production boost from the nudges sent by new users.

We denote the set of newly registered users as N. For each new user $i \in N$, let us assume that the set of existing providers this user chooses to follow as U_i, and the associated set of new following relationships as $E_i := \{(i,u) : u \in U_i\}$. Define $E' := \bigcup_{i \in N} E_i$ as the set of new edges. Then, the additional production boost attributed to the social nudges sent by the new users is given by $\sum_{i\in N}\left(\sum_{e\in E_i}\nu_e\right)$ (Proposition 2 in Online Appendix D.4); it can be reasonably approximated by $\sum_{i\in N}\left(\sum_{e\in E_i}\tilde{\nu}_e(1)\right)$. Hence, the content provider recommendation of each new user can be optimized separately.

For a new user $i \in N$, given the potential content-provider list M_i to recommend, the platform selects $V_i \subset M_i$ with $|V_i| = m$ and recommends the providers in V_i to the new user in a sequential manner. To avoid overly interrupting users, m is generally not too large (i.e., at the magnitude of a few dozen). Denote the probability that a new user will follow the j-th provider recommended to her as c_j, where $c_1 \geq c_2 \geq \cdots \geq c_m$. Let $\pi(j)$ refer to the provider ranked in the j-th position. Then we get the (approximate) additional production boost from the social nudges sent by new user i as $\sum_{j=1}^{m} c_j \tilde{\nu}_{(i,\pi(j))}(1)$. Therefore, the (approximate) "optimal" strategy is to select m providers in M_i with the highest induced (approximate) SNIs and rank them in descending order of induced (approximate) SNI. Similar to optimal seeding, we compare the SNI-based provider recommendation with the benchmark random recommendation, which recommends the content providers based on a random permutation of M_i. By simulation, we also find that the "optimal" strategy significantly outperforms the random strategy in production boost, especially when the recommended-provider-list length m is small. See Online Appendix F.2 for details.

7. Conclusions and Discussion

In two field experiments on a large online content-sharing social network platform, we consistently find that social nudges not only directly boosted nudge recipients' production but also stimulated overall content provision by motivating nudge recipients to send more nudges to others. These effects were amplified when nudge recipients and nudge senders had stronger ties, and they persisted beyond the day nudges were sent.

Inspired by these results, we developed a novel social network model that incorporates the diffusion and over-time effects of social nudges into the estimation of their global effect. We find that the naïve approach simply based on experiments underestimates social nudges' total production boost but our model helps address this issue. Moreover, via simulation examples, we demonstrate that another advantage of adopting our social network model is to find strategies to optimize platform operations regarding social nudges.

Our research offers important practical implications for content-sharing social network platforms. First, social nudges can be a cost-effective intervention for these platforms to lift production on the supply side and consequently increase consumption on the demand side. Platforms are naturally eager to control costs. Compared to financial incentives, social nudges require minimal costs on the platform's end. In fact, due to the success of social nudges observed in our experiments, after the second experiment, Platform O scaled up this function, enabling all users to receive and send social nudges as long as they (or the target they want to nudge) have not uploaded any video for a day or more.

As we noted in Section 2, prior research suggests that peer recognition may not enhance production and could even harm motivation if people do not view the recognized activity as core work in a given setting, doubt the credibility of peer recognition, or see themselves as not qualified for the recognition (Restivo and van de Rijt 2014, Gallus et al. 2020). Those are not concerns in our empirical context. For one thing, providing content is providers' core activity on the platform, and viewers naturally hold the authority to judge providers' content. Thus, recognition from viewers is meaningful to providers. For another thing, since all social nudges on Platform O are spontaneously initiated by followers (rather than being imposed by researchers on providers who might not believe their own qualifications, as in Restivo and van de Rijt 2014), providers who receive social nudges may naturally feel qualified for this form of recognition. In fact, we find that receiving social nudges boosted production among providers with different levels of productivity, including providers who were not very prolific (Section 4.4; Online Appendix C.4). We are hopeful that on content-sharing platforms, nudges from social neighbors could avoid the pitfalls of peer recognition observed in previous research and instead boost production across a broad set of providers.

Second, this work highlights the value of leveraging co-users' influence. Content-sharing social network platforms connect users and facilitate transactions or relationships between users and thus have the advantage of influencing users through interactions between social neighbors, though they have limited power to directly control its providers to produce more content. Thus, platforms can guide co-users to influence each other as a way to improve overall user engagement on platforms.

Likes and positive comments are a prevalent form of co-user influence that may also boost production on a content-sharing platform, but they differ from social nudges in two aspects. One is that while viewers send social nudges because they intentionally want to encourage providers to produce more content, viewers who leave likes or positive comments do not necessarily intend to *actively* influence providers to produce more, and even if they do, their intentions are not clearly conveyed by likes and comments. The other difference is that social nudges are sent by social neighbors, which is not necessarily the case for likes and comments on many content-sharing social network platforms. Prior research has shown that social neighbors are powerful in changing people's behaviors (Bapna and Umyarov 2015, Wang et al. 2018). In our experiments, we also find that stronger ties between neighbors strengthen the effect of social nudges on production, which suggests that the power of social relationships may contribute to the success of social nudges.

Considering these distinctions between likes/comments and social nudges, we speculate that viewers use social nudges differently than likes and positive comments and that social nudges may work on top of likes/comments. As suggestive evidence for our speculation, an additional analysis reveals that sending nudges to providers did not decrease viewers' use of likes and comments (see Online Appendix C.3); and as shown in Section 4.4, social nudges boosted production beyond the first reception day, even when we controlled for the increased likes and comments received by providers, which suggests that providers are motivated by social nudges beyond the influence of likes and comments.

Third, by showcasing that the diffusion of social nudges is crucial for measuring and optimizing the effects of social nudges on production, our work reveals how important it is for platforms to consider the diffusion of an intervention when they decide whether to scale up the intervention and how to maximize its effectiveness. Furthermore, by exploring strategies to maximize the global effect of social nudges—including the optimal seeding strategy and the optimal provider-recommendation strategy for new users—our method may inspire platform managers to leverage a model such as ours to enhance the power of an intervention.

The limitations of our research open up interesting avenues for future research. For one, the type of social nudge we examined is simple, private, and subtle. It was standardized across users, contained simple content, and leveraged no additional psychological principles. It was visible only to recipients in the message center. And as more messages arrived in the message center, earlier social-nudge messages were pushed down, often off the front page of the message center, and become less visible. Using such a light-touch, bare-bones social nudge allows us to provide a clean test of the effect of being nudged, but future research could examine how to design social nudges to produce stronger, longer-lasting effects—for example, by incorporating persuasion techniques and additional psychological insights into nudge messages, allowing senders to write personalized messages, or displaying social nudges publicly in a dedicated area. Another limitation of our research is that we could not causally study the effects of repeatedly receiving social nudges, since the number of social nudges sent to each provider was not exogenous. Future research could randomly assign people to receive varying numbers of nudges and causally estimate their various effects based on the number of nudges received.

Acknowledgments

The authors thank the Department Editor Prof. Victor Martínez-de-Albéniz, the anonymous associate editor, and four referees for their very helpful and constructive comments, which have led to significant improvements in both the content and exposition of this study. We also thank the industry partner for their support on sharing the data and conducting the experiment. Hengchen Dai thanks UCLA Hellman Fellowship and UCLA Faculty Development Award for funding support. Renyu Zhang is grateful for the financial support from the Hong Kong Research Grants Council [16505418] and the Shanghai Eastern Scholar Program [QD2018053].

References

Ashraf, Nava, Oriana Bandiera, B Kelsey Jack. 2014a. No margin, no mission? A field experiment on incentives for public service delivery. *Journal of Public Economics* **120** 1–17.

Ashraf, Nava, Oriana Bandiera, Scott S Lee. 2014b. Awards unbundled: Evidence from a natural field experiment. *Journal of Economic Behavior & Organization* **100** 44–63.

Bai, Jiaru, Kut C So, Christopher S Tang, Xiqun Chen, Hai Wang. 2019. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management* **21**(3) 556–570.

Ballester, Coralio, Antoni Calvó-Armengol, Yves Zenou. 2006. Who's who in networks. wanted: The key player. *Econometrica* **74**(5) 1403–1417.

Banerjee, Siddhartha, Sujay Sanghavi, Sanjay Shakkottai. 2016. Online collaborative filtering on graphs. *Operations Research* **64**(3) 756–769.

Banya, Bockarie Sama. 2017. *The Relationship Between Simple Employee Recognition and Employee Productivity in Business Organizations. A Case Study*. Anchor Academic Publishing.

Bapna, Ravi, Akhmed Umyarov. 2015. Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science* **61**(8) 1902–1920.

Bimpikis, Kostas, Ozan Candogan, Daniela Saban. 2019. Spatial pricing in ride-sharing networks. *Operations Research* **67**(3) 744–769.

Bimpikis, Kostas, Asuman Ozdaglar, Ercan Yildiz. 2016. Competitive targeted advertising over networks. *Operations Research* **64**(3) 705–720.

Bradler, Christiane, Robert Dur, Susanne Neckermann, Arjan Non. 2016. Employee recognition and performance: A field experiment. *Management Science* **62**(11) 3085–3099.

Bramoullé, Yann, Habiba Djebbari, Bernard Fortin. 2020. Peer effects in networks: A survey. *Annual Review of Economics* **12** 603–629.

Buell, Ryan W, Tami Kim, Chia-Jung Tsay. 2017. Creating reciprocal value through operational transparency. *Management Science* **63**(6) 1673–1695.

Buell, Ryan W, Michael I Norton. 2011. The labor illusion: How operational transparency increases perceived value. *Management Science* **57**(9) 1564–1579.

Burtch, Gordon, Qinglai He, Yili Hong, Dokyun Lee. 2021. How do peer awards motivate creative content? Experimental evidence from reddit. *Management Science* Ahead of Print.

Burtch, Gordon, Yili Hong, Ravi Bapna, Vladas Griskevicius. 2018. Stimulating online reviews by combining financial incentives and social norms. *Management Science* **64**(5) 2065–2082.

Cabral, Luis, Lingfang Li. 2015. A dollar for your thoughts: Feedback-conditional rebates on ebay. *Management Science* **61**(9) 2052–2063.

Cachon, Gerard P, Kaitlin M Daniels, Ruben Lobel. 2017. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* **19**(3) 368–384.

Candogan, Ozan, Kostas Bimpikis, Asuman Ozdaglar. 2012. Optimal pricing in networks with externalities. *Operations Research* **60**(4) 883–905.

Candogan, Ozan, Kimon Drakopoulos. 2020. Optimal signaling of content accuracy: Engagement vs. misinformation. *Operations Research* **68**(2) 497–515.

Caro, Felipe, Victor Martínez-de Albéniz. 2020. Managing online content to build a follower base: Model and applications. *INFORMS Journal on Optimization* **2**(1) 57–77.

Celhay, Pablo A, Paul J Gertler, Paula Giovagnoli, Christel Vermeersch. 2019. Long-run effects of temporary incentives on medical care productivity. *American Economic Journal: Applied Economics* **11**(3) 92–127.

Chen, Yan, F Maxwell Harper, Joseph Konstan, Sherry Xin Li. 2010. Social comparisons and contributions to online communities: A field experiment on movielens. *American Economic Review* **100**(4) 1358–98.

Chen, Yubo, Qi Wang, Jinhong Xie. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research* **48**(2) 238–254.

Cohen, Maxime C, Pavithra Harsha. 2020. Designing price incentives in a network with social interactions. *Manufacturing & Service Operations Management* **22**(2) 292–309.

Cui, Ruomeng, Jun Li, Dennis J Zhang. 2020a. Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on airbnb. *Management Science* **66**(3) 1071–1094.

Cui, Ruomeng, Meng Li, Qiang Li. 2020b. Value of high-quality logistics: Evidence from a clash between sf express and alibaba. *Management Science* **66**(9) 3879–3902.

Cui, Ruomeng, Dennis J Zhang, Achal Bassamboo. 2019. Learning from inventory availability information: Evidence from field experiments on amazon. *Management Science* **65**(3) 1216–1235.

Dai, Hengchen, Zhiyu Zeng, Dennis Zhang, Zhiwei Xu, Zuo-Jun Max Shen. 2022. The value of customer-related information on service platforms: Evidence from a large field experiment. *Available at SSRN 3528619* .

De Grip, Andries, Jan Sauermann. 2012. The effects of training on own and co-worker productivity: Evidence from a field experiment. *The Economic Journal* **122**(560) 376–399.

Dukas, Reuven. 2004. Causes and consequences of limited attention. *Brain, Behavior and Evolution* **63**(4) 197–210.

Frey, Bruno, Jana Gallus. 2017. *Honours versus Money: The Economics of Awards*. Oxford University Press.

Gallus, Jana. 2017. Fostering public good contributions with symbolic awards: A large-scale natural field experiment at wikipedia. *Management Science* **63**(12) 3999–4015.

Gallus, Jana, Olivia Jung, Karim R Lakhani. 2020. Recognition incentives for internal crowdsourcing: A field experiment at nasa. *Harvard Business School Technology & Operations Management Unit Working Paper* (20-059).

Gee, Laura K. 2019. The more you know: Information effects on job application rates in a large field experiment. *Management Science* **65**(5) 2077–2094.

Gelper, Sarah, Ralf van der Lans, Gerrit van Bruggen. 2021. Competition for attention in online social networks: Implications for seeding strategies. *Management Science* **67**(2) 1026–1047.

Goes, Paulo B, Chenhui Guo, Mingfeng Lin. 2016. Do incentive hierarchies induce user effort? Evidence from an online knowledge exchange. *Information Systems Research* **27**(3) 497–516.

Grant, Adam M, Francesca Gino. 2010. A little thanks goes a long way: Explaining why gratitude expressions motivate prosocial behavior. *Journal of Personality and Social Psychology* **98**(6) 946.

Gurvich, Itai, Martin Lariviere, Antonio Moreno. 2019. Operations in the on-demand economy: Staffing services with self-scheduling capacity. *Sharing Economy*. Springer, 249–278.

Horn, Roger A, Charles R Johnson. 2012. *Matrix Analysis*. Cambridge University Press.

Huang, Ni, Gordon Burtch, Bin Gu, Yili Hong, Chen Liang, Kanliang Wang, Dongpu Fu, Bo Yang. 2019. Motivating user-generated content with performance feedback: Evidence from randomized field experiments. *Management Science* **65**(1) 327–345.

Iacus, Stefano M, Gary King, Giuseppe Porro. 2012. Causal inference without balance checking: Coarsened exact matching. *Political Analysis* **20**(1) 1–24.

Jackson, Matthew O. 2005. The economics of social networks .

Jackson, Matthew O. 2010. *Social and Economic Networks*. Princeton University Press.

Kabra, Ashish, Elena Belavina, Karan Girotra. 2020. Bike-share systems: Accessibility and availability. *Management Science* **66**(9) 3803–3824.

Konings, Jozef, Stijn Vanormelingen. 2015. The impact of training on productivity and wages: Firm-level evidence. *Review of Economics and Statistics* **97**(2) 485–497.

Kosfeld, Michael, Susanne Neckermann. 2011. Getting more work for nothing? Symbolic awards and worker performance. *American Economic Journal: Microeconomics* **3**(3) 86–99.

Kuang, Lini, Ni Huang, Yili Hong, Zhijun Yan. 2019. Spillover effects of financial incentives on non-incentivized user engagement: Evidence from an online knowledge exchange platform. *Journal of Management Information Systems* **36**(1) 289–320.

Kuhn, Peter, Peter Kooreman, Adriaan Soetevent, Arie Kapteyn. 2011. The effects of lottery prizes on winners and their neighbors: Evidence from the dutch postcode lottery. *American Economic Review* **101**(5) 2226–47.

Lazear, Edward P. 2000. Performance pay and productivity. *American Economic Review* **90**(5) 1346–1361.

Li, Yuanchen, Lauren Xiaoyuan Lu, Susan F Lu. 2020. Do social media trump government report cards in influencing consumer choice? Evidence from us nursing homes. *Evidence from US Nursing Homes (January 24, 2020)* .

Mas, Alexandre, Enrico Moretti. 2009. Peers at work. *American Economic Review* **99**(1) 112–45.

Mohan, Bhavya, Ryan W Buell, Leslie K John. 2020. Lifting the veil: The benefits of cost transparency. *Marketing Science* **39**(6) 1105–1121.

Mookerjee, Radha, Subodha Kumar, Vijay S Mookerjee. 2017. Optimizing performance-based internet advertisement campaigns. *Operations Research* **65**(1) 38–54.

Moon, Ken, Patrick Bergemann, Daniel Brown, Andrew Chen, James Chu, Ellen Eisen, Gregory Fischer, Prashant Kumar Loyalka, Sungmin Rho, Joshua Cohen. 2018. Manufacturing productivity with worker turnover. *Available at SSRN 3248075* .

Nicoletti, Cheti, Kjell G Salvanes, Emma Tominey. 2018. The family peer effect on mothers' labor supply. *American Economic Journal: Applied Economics* **10**(3) 206–34.

Papanastasiou, Yiangos, Nicos Savva. 2017. Dynamic pricing in the presence of social learning and strategic consumers. *Management Science* **63**(4) 919–939.

Parker, Chris, Kamalini Ramdas, Nicos Savva. 2016. Is it enough? Evidence from a natural experiment in India's agriculture markets. *Management Science* **62**(9) 2481–2503.

Pew Research Center. 2010. Online product research. (september 29). URL https://www.pewresearch.org/internet/2010/09/29/online-product-research-2/.

Restivo, Michael, Arnout van de Rijt. 2014. No praise without effort: Experimental evidence on how rewards affect wikipedia's contributor community. *Information, Communication & Society* **17**(4) 451–462.

Roels, Guillaume, Xuanming Su. 2014. Optimal design of social comparison effects: Setting reference groups and reference points. *Management Science* **60**(3) 606–627.

Ryan, Richard M, Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist* **55**(1) 68.

Sacerdote, Bruce. 2001. Peer effects with random assignment: Results for dartmouth roommates. *The Quarterly Journal of Economics* **116**(2) 681–704.

Song, Hummy, Anita L Tucker, Karen L Murrell, David R Vinson. 2018. Closing the productivity gap: Improving worker productivity through public relative performance feedback and validation of best practices. *Management Science* **64**(6) 2628–2649.

Tan, Tom Fangyun, Serguei Netessine. 2014. When does the devil make work? An empirical study of the impact of workload on worker productivity. *Management Science* **60**(6) 1574–1593.

Tan, Tom Fangyun, Serguei Netessine. 2019. When you work with a superman, will you also fly? An empirical study of the impact of coworkers on performance. *Management Science* **65**(8) 3495–3517.

Tan, Tom Fangyun, Serguei Netessine. 2020. At your service on the table: Impact of tabletop technology on restaurant performance. *Management Science* **66**(10) 4496–4515.

Thaler, Richard H, Cass R Sunstein. 2009. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. Penguin.

Wang, Chong, Xiaoquan Zhang, Il-Horn Hann. 2018. Socially nudged: A quasi-experimental study of friends' social influence in online product ratings. *Information Systems Research* **29**(3) 641–655.

Whitmore, Diane. 2005. Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review* **95**(2) 199–203.

Xu, Yuqian, Mor Armony, Anindya Ghose. 2021. The interplay between online reviews and physician demand: An empirical investigation. *Management Science* **67**(12) 7344–7361.

Yang, Jiang, Xiao Wei, Mark Ackerman, Lada Adamic. 2010. Activity lifespan: An analysis of user survival patterns in online knowledge sharing communities. *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 4.

Zhang, Dennis J, Gad Allon, Jan A Van Mieghem. 2017. Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses. *Manufacturing & Service Operations Management* **19**(3) 347–367.

Zhang, Dennis J, Hengchen Dai, Lingxiu Dong, Fangfang Qi, Nannan Zhang, Xiaofei Liu, Zhongyi Liu, Jiang Yang. 2020. The long-term and spillover effects of price promotions on retailing platforms: Evidence from a large randomized experiment on alibaba. *Management Science* **66**(6) 2589–2609.

Zhang, Dennis J, Hengchen Dai, Lingxiu Dong, Qian Wu, Lifan Guo, Xiaofei Liu. 2019. The value of pop-up stores on retailing platforms: Evidence from a field experiment with Alibaba. *Management Science* **65**(11) 5142–5151.

Zhou, Junjie, Ying-Ju Chen. 2016. Targeted information release in social networks. *Operations Research* **64**(3) 721–735.

Online Appendices

A. Robustness Checks of the Main Results From the First Social-Nudge Experiment

A.1. Analyzing All Providers Who Were Sent at Least One Social Nudge in the Experiment

For analyses reported in the main text, we focused on providers who had never received any social nudges before the first social nudge experiment (as explained in Section 3), in order to estimate how social nudges change behavior when a platform starts to implement the social nudge function. In this section, we report the production-boosting and diffusion effects of social nudges among all providers whose followers sent them at least one social nudge during our experiment (N $=$ $1,946,118$), as a robustness check.

Using regression specification (1), we predicted the number of videos a provider uploaded (*Number of Videos Uploaded*) and the number of social nudges sent by a provider to other providers (*Number of Social Nudges Sent*) on the first reception day. As shown in Table 10, receiving social nudges boosted the number of videos upload on the first reception day by 9.53% (0.0222 standard deviations; $p < 0.0001$), and increased the number of social nudges sent to other providers by 13.92% (0.0323 standard deviations; $p < 0.0001$). Therefore, the immediate effects of receiving social nudges are qualitatively unchanged if we examine all providers whose followers sent them at least one social nudge during our experiment.

Table 10 Effects of Social Nudges Among All Providers Who Were Sent at Least One Social Nudge in the First Social-Nudge Experiment

Outcome Variable	Number of Videos Uploaded	Number of Social Nudges Sent
	on the First Reception Day	
	(1)	(2)
Treatment	0.0222****	0.0323****
	(0.0014)	(0.0014)
Relative Effect Size	9.53%	13.92%
Observations	$1,946,118$	$1,946,118$

Note: Number of Videos Uploaded and Number of Social Nudges Sent were standardized to have a unit standard deviation before entering the regressions. Columns (1)–(2) include all providers whose followers sent them at least one social nudge during our first social-nudge experiment. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

A.2. Predicting Production Within 24 Hours Following the First Nudge

In the main text, we examined videos providers uploaded on the first reception day, defined as the calendar date the first social nudge was sent to them during the experiment. As a robustness

Table 11 Effects of Social Nudges on Content Production Within 24 Hours Following the First Nudge

Outcome Variable	Number of Videos Uploaded Within 24 Hours Following the First Nudge	Upload Incidence Within 24 Hours Following the First Nudge
	(1)	(2)
Treatment	0.0297****	0.0131****
	(0.0020)	(0.0006)
Relative Effect Size	12.45%	13.22%
Observations	993,676	993,676

Notes: Number of Videos Uploaded Within 24 Hours Following the First Nudge was standardized to have a unit standard deviation before entering the regression. Columns (1)–(2) include all providers in the sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

check, we tracked the number of videos each provider uploaded during 24 hours since the first social nudge. We constructed two outcome variables: *Number of Videos Uploaded Within 24 Hours Following the First Nudge* and *Upload Incidence Within 24 Hours Following the First Nudge*. We predicted these outcome variables using regression specification (1).

As shown in Table 11, treatment providers boosted the number of videos uploaded within 24 hours following the first social nudge by 12.45% (0.0297 standard deviations; p<0.0001; column (1)) and increased their likelihood of uploading any video within 24 hours following the first social nudge by 13.22% (1.31 percentage points; p<0.0001; column (2)). These results suggest that the positive effect of social nudges on content production is robust in terms of this alternative time frame.

B. The Second Social-Nudge Experiment as a Replication

We conducted another experiment to replicate the effects of social nudges on production and diffusion observed in the first field experiment (Sections 4 and 5). The replication experiment lasted from 5pm on September 14, 2018 to the end of September 20, 2018. It lasted longer than the main experiment and targeted a nonoverlapping but smaller sample of providers than the main experiment.[1] Providers targeted by the replication experiment were randomly assigned into either the treatment condition or the control condition. Similar to our main experiment (Section 3), our analyses of the replication experiment focused on providers who satisfied two criteria: (1) at least one of their followers sent them a social nudge during the experimental period, and (2) they had never received any social nudges before the experiment. Our final analysis sample consisted of 678,090 qualified providers, among whom 338,415 were in the treatment condition and 339,675 were in the control condition.

[1] We first randomly sampled a portion of providers to be included in the main experiment. Then among the remaining providers, we randomly sampled a smaller portion of providers to be involved in the replication experiment.

B.1. Direct Effects of Social Nudges on Content Production (Replicated)

We first examined the production boosting effect of receiving social nudges over time in the second experiment. Specifically, using regression specification (1), we predicted the number of videos uploaded each day from the first reception day on until the first day when the difference between two conditions was not statistically significant. We report the estimation results in Table 12, which shows that the effect sizes observed in the second experiment are comparable to the effect sizes observed in the main experiment (Table 4). Therefore, our results on the direct production-boosting effect of social nudges are robust.

Table 12 Over-Time Direct Effects of Social Nudges on Content Production (Replicated)

Outcome Variable	Number of Videos Uploaded			
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4
	(1)	(2)	(3)	(4)
Treatment	0.0228****	0.0107****	0.0083***	0.0033
	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Relative Effect Size	11.83%	7.79%	3.96%	
Observations	678,090	678,090	678,090	678,090

Notes: Number of Videos Uploaded was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on Day t relative to the first reception day, where $t = 1$ means the first reception day. Columns (1)–(4) include all providers in our sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

B.2. Effects of Social Nudges on Nudge Diffusion (Replicated)

Next we tested the diffusion effect of receiving social nudges over time in the second experiment. Specifically, for each day t starting from the first reception day, we predicted the number of social nudges sent on that day using regression specification (1) until the first day when the difference between two conditions was not statistically significant. We report the estimation results in Table 13, which shows that the effect sizes observed in the second experiment are comparable to the effect sizes observed in the main experiment (Table 6). Therefore, our results on the diffusion of social nudges are robust.

C. Additional Analyses about the Direct Effects of Social Nudges

C.1. Addressing an Alternative Explanation about Control Providers' Resentment

As explicated in Section 1, we expected social nudges to boost content production because providers receiving social nudges might feel more valued by others and thus more motivated to supply effort. However, one potential alternative explanation for our observed difference in video production between treatment and control providers is that through other ways beyond the message center, control providers realized that their followers sent them social nudges but they could not receive

Table 13 Over-Time Effects of Social Nudges on Nudge Diffusion (Replicated)

Outcome Variable	Number of Social Nudges Sent				
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4	On Day 5
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0325****	0.0215****	0.0084***	0.0057*	0.0039
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
Relative Effect Size	16.25%	14.16%	5.78%	4.02%	
Observations	678,090	678,090	678,090	678,090	678,090

Notes: Number of Social Nudges Sent was standardized to have a unit standard deviation before entering the regressions. The unit of analysis for all columns was a provider on Day t relative to the first reception day, where $t = 1$ refers to the first reception day. Columns (1)–(5) include all providers in our sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

these social nudges, which might make control providers feel resentful toward the platform and thus reduce their production. As mentioned in Section 3, the only way for users to directly communicate with each other on the platform is through the private-message function. It is plausible that during our experiment followers privately messaged providers after sending them social nudges, which led control providers to realize that they were blocked from viewing nudges. To address this alternative explanation, we conducted two sets of additional analyses.

First, we examined how private messages influenced control providers. If control providers knew via private messages that their followers sent them a social nudge but they were not allowed to see the nudge and if this created resentment, we should expect that receiving private messages from followers who sent them social nudges during the experiment negatively impacted control providers' content production. To test this possibility, we used the DiD method. This analysis included two observations per control provider, with one observation corresponding to the first reception day and one observation corresponding to the day before the experiment. For each observation of provider i, her content production equaled the number of videos uploaded on the corresponding day (either the first reception day or the day before the experiment). The DiD regression specification is formulated as below

$$\begin{aligned} Outcome\ Variable_{it} = \beta_0 + \beta_1 Private\ Messages\ Incidence_i + \beta_2 First\ Reception\ Day_{it} \\ + \beta_3 Private\ Messages\ Incidence_i * First\ Reception\ Day_{it} + \epsilon_{it} \end{aligned} \quad (12)$$

whereby $Private\ Messages\ Incidence_i$ was a binary variable that equaled one if the follower who sent provider i the first social nudge in the experiment (i.e., provider i's first social-nudge sender) also sent any private messages to i between the start date of the experiment and provider i's first reception day (including both ends) and zero otherwise[2]; and $First\ Reception\ Day_{it}$ was a binary

[2] To protect user privacy, Platform O could not share the content of private messages with us. Thus, we could not use content analysis to identify whether each provider's first social-nudge sender told the provider about the nudge in their private communications, but instead we used whether a provider received private messages from their first social-nudge sender as a proxy, since receiving such private messages was the only plausible channel for control providers to find out the blocking of social nudges.

Table 14 The Role of Private Messages in Content Production Among Control Providers

Panel A: DiD Analysis about Private Messages Among Control Providers

Outcome Variable	Number of Videos Uploaded
	(1)
Private Messages Incidence	0.1603^{****}
	(0.0116)
First Reception Day	0.1258^{****}
	(0.0020)
Private Messages Incidence * First Reception Day	0.2248^{****} (0.0203)
Observations	993,400

Panel B: Comparison of Two Subsamples Based on Private Message Incidence

Outcome Variable	Number of Videos Uploaded	
Subsample	*Providers Who Received Any Private Messages From the First Social-Nudge Sender*	*Providers Who Received No Private Messages From the First Social-Nudge Sender*
	(1)	(2)
Treatment	0.0710^{***}	0.0235^{****}
	(0.0192)	(0.0020)
Relative Effect Size	14.15%	12.36%
Observations	28,142	965,534

Notes: Number of Videos Uploaded was standardized to have a unit standard deviation before entering the regressions. Panel A includes all control providers in our sample, with each control provider contributing two observations. Standard errors in Panel A are clustered at the provider level. Column (1) in Panel B includes treatment and control providers who received any private messages from their first social-nudge sender between the start date of the experiment and the first reception day, and column (2) in Panel B includes treatment and control providers who did not receive any private messages from their first social-nudge sender in this period. Robust standard errors reported in the parentheses in Panel B. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

variable that equaled one if an observation corresponded to the first reception day and zero if the observation corresponded to the day before the experiment. We clustered standard errors by provider.

As shown in Table 14 Panel A, since the coefficient on the interaction between *Private Messages Incidence*$_i$ and *First Reception Day*$_{it}$ is positive ($p < 0.0001$), we have no evidence to suggest that receiving private messages from followers who sent them social nudges during the experiment would reduce control providers' content production.

Second, we split the whole provider sample in our experiment into two subsamples based on whether each provider's first social-nudge sender sent any private messages to the provider between the start date of the experiment and provider i's first reception day (including both ends). Within each subsample, we compared the *Number of Videos Uploaded* on the first reception day between treatment and control conditions using regression specification (1).

As shown in Table 14 Panel B, no matter whether a provider got any private messages from their first social-nudge sender, receiving social nudges increased treatment providers' content production, relative to control providers' (both p-values < 0.001). The relative effect size is very similar among providers who got private messages from their first social-nudge sender (14.15% as shown in column (1)) and among providers who did not get private messages from their first social-nudge sender (12.36% as shown in column (2)).

Altogether, these results do not support the alternative explanation: it is unlikely that communication from followers via private messages led control providers to find out they could not view social nudges, elicited resentment, and thus reduced their motivation to produce videos. In addition, note that all broadcasters selected into our analysis sample had not received any social nudges before the experiment. Thus, it is unlikely for providers in our analysis to naturally realize that they did not receive social nudges during the experiment without any hints from followers.

C.2. Role of Likes and Comments

As mentioned in Section 3, when viewers watch a video, they can mark that they like the video and leave comments below the video. Since receiving social nudges could immediately boost video production (see Section 4.1), nudge recipients may also immediately receive more likes and comments due to the increased number of videos uploaded. Such positive feedback from viewers may in turn motivate nudge recipients to produce more videos going forward. This raises the question of whether and to what extent the immediate increase in likes and comments due to receiving social nudges drives the observed over-time effect of social nudges on nudge recipients' content production (as shown in Section 4.3).

To answer this question, we first tested whether receiving social nudges led the recipient to obtain more likes and comments. For each provider on her first reception day, we calculated the number of likes and comments she obtained that day (*Number of Likes on the First Reception Day* and *Number of Comments on the First Reception Day*, respectively). We winsorzied these two variables at the 95th percentile of their respective nonzero values because they were highly skewed (due to a small number of providers being too popular). Using regression specification (1), we predicted these two outcome variables. As shown in Table 15 Panel A, treatment providers obtained more likes than control providers on the first reception day by 0.0112 standard deviations, or 4.52% ($p < 0.0001$; column (1)); treatment providers obtained more comments than control providers on the first reception day by 0.0108 standard deviations, or 5.40% ($p < 0.0001$; column (2)). Hence, treatment providers obtained more likes and comments after they received social nudges, relative to control providers.

We next tested how much the immediate increase in likes and comments due to the receipt of social nudges contributed to the effects of receiving social nudges on content production during the few days after the first reception day. In one series of regressions, we predicted the number of videos uploaded the day following the first reception day using regression specification (1), and we

Table 15 Effects of Social Nudges on Content Production With or Without Controlling for the Role of Likes and Comments

Panel A

Outcome Variable	Number of Likes on the First Reception Day	Number of Comments on the First Reception Day
	(1)	(2)
Treatment	0.0112****	0.0108****
	(0.0020)	(0.0020)
Relative Effect Size	4.52%	5.40%
Observations	993,676	993,676

Panel B

Outcome Variable	Number of Videos Uploaded Following the First Reception Day			
	(1)	(2)	(3)	(4)
Treatment	0.0129****	0.0091****	0.0094****	0.0090****
	(0.0020)	(0.0019)	(0.0019)	(0.0019)
Number of Likes on the First Reception Day		0.3374****		0.2299****
		(0.0009)		(0.0018)
Number of Comments on the First Reception Day			0.3225****	0.1253****
			(0.0009)	(0.0018)
Relative Effect Size	5.29%	3.74%	3.87%	3.68%
Observations	993,676	993,676	993,676	993,676

Panel C

Outcome Variable	Number of Videos Uploaded on the Second Day Following the First Reception Day			
	(1)	(2)	(3)	(4)
Treatment	0.0065**	0.0061**	0.0063**	0.0061**
	(0.0020)	(0.0020)	(0.0020)	(0.0020)
Number of Likes the Day Following the First Reception Day		0.1209****		0.0853****
		(0.0010)		(0.0020)
Number of Comments the Day Following the First Reception Day			0.1152****	0.0408****
			(0.0010)	(0.0020)
Relative Effect Size	2.54%	2.40%	2.46%	2.42%
Observations	993,676	993,676	993,676	993,676

Notes: All continuous variables were standardized to have a unit standard deviation before entering the regressions. All columns in Panels A, B, and C include all providers in the sample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

compared the regression results with or without controlling for the number of likes or comments a provider obtained the day before (i.e., *Number of Likes on the First Reception Day* and *Number of Comments on the First Reception Day*). According to Table 15 Panel B, without controlling for likes or comments on the first reception day, we find that receiving social nudges boosted the number of videos uploaded by 0.0129 standard deviations (5.29%, $p < 0.0001$; column (1) of Panel A) on the day following the first reception day. This effect reduced but remained statistically significantly when we controlled for the number of likes a provider got on her first reception day (column (2)), the number of comments she got on her first reception day (column (3)), or both (column (4)).

Further, we predicted the *number of videos uploaded* on the *second day* following the first reception day using regression specification (1), and we compared the regression results with or without controlling for the number of likes or comments a provider obtained the day before (i.e., on the day following the first reception day). As shown in Table 15 Panel C, the positive effect of receiving social nudges on content production two days later reduced only slightly when we added these control variables.

Altogether, these findings indicate that getting more likes and comments after treatment providers uploaded more videos in response to social nudges contributed to some extent to the over-time effect of social nudges on content production. However, increased likes and comments are not the only reason, neither are the primary reason why the effect of receiving social nudges on content production lasted for days, since we observe only a slight to moderate decrease in the magnitude of the production-boosting effect of social nudges after the first reception day when we controlled for likes and comments providers obtained earlier. This suggests that receiving social nudges per se is sufficient to motivate video production a few days, even without additional positive feedback providers receive due to their increased production along the way[3].

C.3. Do Social Nudges Cannibalize Likes/Comments?

Sending social nudges is a new way for viewers to interact with providers and express appreciation of their videos on top of the existing mechanisms including likes and (positive) comments. A natural question is whether viewers will mark fewer likes and leave fewer comments once they begin to use social nudges. In other words, will social nudges cannibalize the use of likes and comments? We address this question in two aspects. First, from the perspective of providers, we find that receiving social nudges did not lead providers to receive fewer likes and comments; if anything, we causally document that social nudges brought providers more likes and comments in the next couple of days (see Online Appendix C.2). Second, from the perspective of viewers, we tested whether sending social nudges decreased the usage of likes and comments, and we report this analysis in this subsection.

When we investigated this question, Platform O no longer stored detailed data about likes and comments that took place around our experimental period; thus, we leveraged observational data in December 2021. We first randomly sampled $10,000,000$ viewers who logged onto Platform O between December 1, 2021 and December 7, 2021. For each viewer in the sample, we calculated

[3] We conducted these analyses because it is *theoretically* interesting to tease apart whether the lingering effect of receiving social nudges on content production is driven by providers obtaining an increased amount of positive feedback on their videos or by providers feeling motivated by social nudges per se. But *practically speaking*, we believe the feedback mechanism is meaningful because increased likes and comments are *consequences* of the initial boost in content production in response to social nudges. Thus, we do not distinguish these mechanisms when we calculate the global impact of social nudges in Section 6.

Table 16 Users Who Sent Social Nudges Did Not Reduce the Usage of Likes/Comments

Outcome Variable	Number of Likes Marked (1)	Number of Comments Left (2)
Incidence of Sending Social Nudges	0.2594****	0.1836****
	(0.0023)	(0.0023)
Post	-0.0057****	-0.0008
	(0.0006)	(0.0004)
Incidence of Sending Social Nudges * Post	0.0556****	0.0323 ****
	(0.0010)	(0.0010)
Observations	1,412,164	1,412,164

Notes: Both outcome variables are standardized to have a unit deviation. Robust standard errors clustered at the viewer level are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

whether she sent at least one social nudge in that period (*Incidence of Sending Social Nudges*), and obtained a host of features, including gender, age, her residential community type (i.e., countryside, town, or urban area), the tier of city she lives in (e.g., first tier, second tier, etc.), the number of followers she had on November 30, 2021, the number of users she was following on November 30, 2021, whether she had uploaded any video by November 30, 2021, how long she was active on Platform O in the previous week (November 24—November 30, 2021), how long she watched videos on Platform O in the previous week, and whether she was banned from social interactions (such as marking likes and sending comments) on Platform O in the previous week. Via Coarsened Exact Matching (CEM, e.g., Iacus et al. 2012), we constructed two matched groups: the "treatment" group who sent at least one social nudge in the week of December 1—December 7, 2021 ($353,041$ viewers; $Incidence\ of\ Sending\ Social\ Nudges = 1$), and the "control" group who did not send any social nudge that week ($353,041$ viewers; $Incidence\ of\ Sending\ Social\ Nudges = 0$). For each viewer i in the matched sample and for each outcome variable, we measured two observations, with one observation corresponding to the week of December 1—December 7, 2021 ($Post = 1$), and the other observation corresponding to the prior week (November 24—November 30, 2021; $Post = 0$). The DiD regression specification is formulated as

$$\begin{aligned} Outcome\ Variable_{it} = \beta_0 + \beta_1 Incidence\ of\ Sending\ Social\ Nudges_i + \beta_2 Post_{it} \\ + \beta_3 Incidence\ of\ Sending\ Social\ Nudges_i * Post_{it} + \epsilon_{it} \end{aligned} \tag{13}$$

where the number of likes marked by viewer i during week t ($Number\ of\ Likes\ Marked_{it}$) and the number of comments left by viewer i during week t ($Number\ of\ Comments\ Left_{it}$) serve as the outcome variables. We clustered standard errors by viewers.

As shown in Table 16, the coefficient on the interaction term between $Incidence\ of\ Sending\ Social\ Nudges_i$ and $Post_{it}$ is positive and significant for both outcome variables (both p-values < 0.0001). This suggests that viewers who sent any social nudges marked more likes and left more comments, compared to viewers who did not send social nudges. Thus, combining the identification strategies of matching and DiD, we find no evidence for cannibalization of social nudges on likes and comments.

C.4. Effects of Social Nudges Across Providers With Different Baseline Productivity

The scant prior literature that has examined the causal effects of peer recognition without financial incentives has not provided a clear answer to the question of whether peer recognition can boost recipients' production (Restivo and van de Rijt 2014, Gallus et al. 2020). In a field experiment involving top 10% of providers to Wikipedia, Restivo and van de Rijt (2014) found that peer recognition increased only the most productive 1% of content providers but not providers ranked at the 91st–99th percentile. To test whether the production-boosting effect of social nudges can generalize to providers with different levels of baseline productivity, we divided the providers in our sample into three subsamples: providers whose historical production—the number of videos uploaded during the week prior to the experiment—was (1) below or at the 90th percentile of the distribution of historical production across all providers in the sample ("*low-productivity providers*"; ignored by Restivo and van de Rijt (2014)), (2) in the 91st–99th percentile range ("*medium-productivity providers*"; comparable to the definition of less-productive providers in Restivo and van de Rijt (2014)), and (3) at the 100th percentile ("*high-productivity providers*"; comparable to the most productive 1% providers in Restivo and van de Rijt (2014)). For each subsample, we separately estimated the effect of receiving social nudges on the day of nudges being sent. That is, we predicted the number of videos uploaded on the first reception day using regression specification (1).

As shown in Table 17, the number of videos uploaded on the first reception day was boosted by 19.37% among low-productivity providers (0.0220 standard deviations; $p < 0.0001$; column (1)), by 6.91% among medium-productivity providers (0.0577 standard deviations; $p < 0.0001$; column (2)), and by 7.67% among high-productivity providers (0.2145 standard deviations; $p < 0.05$; column (3)). Overall, these results suggest that not only the most productive 1% of providers but also providers whose historical production was in the 0th-99th percentile range were also motivated by receiving social nudges.

C.5. Comparison Between Social Nudges and Platform-Initiated Nudges

Receiving a social nudge and its implied recognition may make providers feel more valued, thus motivating them to produce new videos. However, such information communicated via social nudges from neighbors may not be passed on by nudges from the platform to encourage production. To explore whether social nudges have effects beyond regular nudges sent from companies, we leveraged another randomized field experiment that tested the effects of receiving nudges from

Table 17 Direct Effects of Social Nudge Across Providers With Different Historical Production Levels

Outcome Variable	Number of Videos Uploaded on the First Reception Day		
Subsample	*Low-Productivity Providers*	*Medium-Productivity Providers*	*High-Productivity Providers*
	(1)	(2)	(3)
Treatment	0.0220****	0.0577****	0.2145*
	(0.0015)	(0.0124)	(0.0926)
Relative Effect Size	19.37%	6.91%	7.67%
Observations	901,286	83,838	8,552

Notes: Number of videos uploaded was standardized to have a unit standard deviation before entering the regressions. Each column includes the providers in the corresponding subsample. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

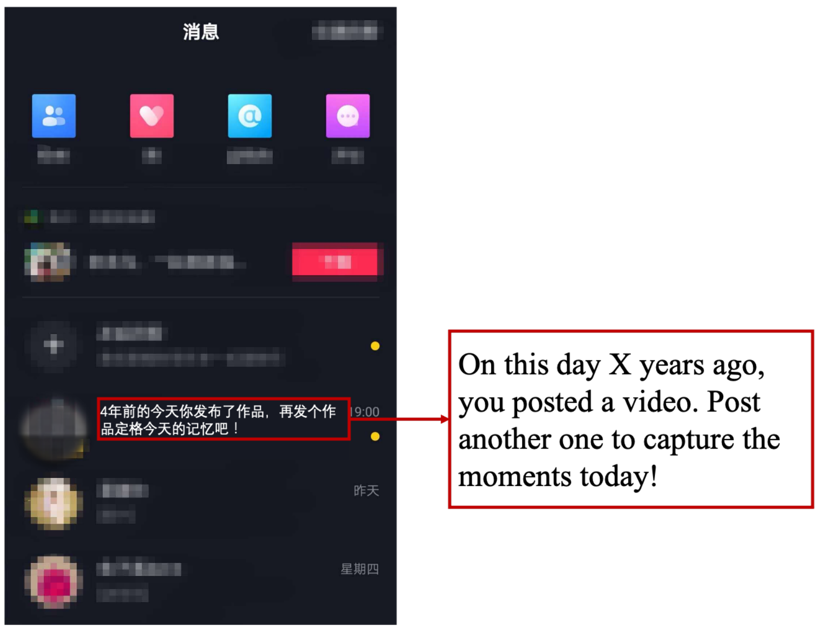


Figure 3 A Platform-Initiated Nudge[4]

Platform O. We refer to this experiment as the *platform-initiated nudge experiment*. The platform-initiated nudge experiment randomly targeted a subset of users on Platform O, no matter whether they were targeted by the social-nudge experiments. We note that comparing the results of our main social-nudge experiment versus the platform-initiated nudge experiment does not causally estimate the difference between social nudges and platform-initiated nudges. Specifically, since the two experiments were conducted in different time periods and providers were not randomly assigned to receive one of these two types of nudges, the providers included in the two experiments were not exactly comparable. As described below, we sought to construct samples from the two experiments that were as comparable to each other as possible.

[4] In order to protect Platform O's identity, similar to how we dealt with Figure 1, we created Figure 3 by modifying the app interface of a widely-used video-sharing platform. Platform O has a similar app interface to Figure 3.

Experiment Design and Data. The platform-initiated nudge experiment was conducted between 9AM on July 22, 2019 and 5AM on August 30, 2019. Half of the providers were randomly assigned to the treatment condition, and the other half to the control condition. During the experiment, the platform identified providers who published a video one or more years ago exactly on the same date. For these providers, Platform O created a message that read, "On this day X years ago, you posted a video. Post another one to capture the moments today!" where "X" was filled in with the actual number of years that had elapsed.[5] The only factor that we manipulated between treatment and control providers was that Platform O actually sent out the aforementioned message to treatment providers on that date, but not to control providers. Therefore, control providers could not receive any platform-initiated nudges. Messages about platform-initiated nudges were displayed in the message center, the same as social nudges (see Figure 1 (b)).

We first selected treatment and control providers who were qualified to be sent at least one message from Platform O during our experiment. For these providers, we defined "the first reception day" as the day when they first became qualified to be sent Platform O's nudge message. The sample selection criteria and the definition of the first reception day here match our approach in the social-nudge experiments.

The Effects of Receiving Platform-Initiated Nudges on Production. We first examined the effects of receiving platform-initiated nudges on content production, both on the first reception day and in the next few days. Consistent with our analytical strategy for the social-nudge experiments (Section 4.3), we examined how receiving nudges from Platform O affected providers' production each day between treatment and control providers among the full sample of providers from the first reception day on until the first day when the difference between conditions was not statistically significant. Specifically, for each day t starting from the first reception day (where t equals $1, 2, \cdots$ and $\mathrm{t} = 1$ refers to the first reception day itself), we predicted the number of videos uploaded on that day using regression specification (1).

We report the regression results in Table 18 Panel A. On the first reception day, the platform-initiated nudge treatment lifted the number of videos uploaded by 0.0105 standard deviations ($p < 0.0001$), which amounts to a 5.55% increase relative to the average in the control condition, as shown in column (1). On Day 2 (the day following the first reception day), the number of video uploaded was higher in the treatment condition than in the control condition by 0.0026 standard deviations, or 1.37% ($p < 0.0001$; column (2)); on Day 3 (the second day from the first reception

[5] To avoid disturbing providers, Platform O sent out a maximum of two messages to each provider in one week. Specifically, on the first day of each week during the experiment, for provider i, Platform O identified the dates during that week on which provider i uploaded any video exactly one or more years ago. If more than two dates satisfied the criterion, Platform O picked the dates on which the video uploaded exactly one or more years ago had the highest or second highest views (among all videos uploaded in the same week one or more years ago).

Table 18 Comparison of Social Nudges and Platform-Initiated Nudges

Panel A: Direct Effects of Platform-Initiated Nudges on Content Production

Outcome Variable	Number of Videos Uploaded			
	On Day 1 (First Reception Day)	On Day 2	On Day 3	On Day 4
	(1)	(2)	(3)	(4)
Treatment	0.0105****	0.0026****	0.0019**	0.0011
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Relative Effect Size	5.55%	1.37%	0.99%	
Observations	11,043,476	11,043,476	11,043,476	11,043,476

Panel B: Comparison of Social Nudges and Platform-Initiated Nudges Using an Overlapping Sample of Providers

Outcome Variable	Number of Videos Uploaded On Day 1 (First Reception Day)	
	Platform-Initiated Nudges	*Social Nudges*
Treatment	0.0152	0.0216***
	(0.0093)	(0.0063)
Relative Effect Size		12.35%
Observations	63,467	63,467

Notes: Number of videos uploaded was standardized to have a unit standard deviation before entering the regressions. Panel A includes all providers who satisfied sample selection criteria for the platform-initiated nudge experiment. The unit of analysis in Panel A was a provider on Day t relative to the first reception day, where $t = 1$ means the first reception day. Panel B includes providers who were selected for both the social-nudge experiment and the platform-initiated nudge experiment. The unit of analysis in Panel B was a provider on her first reception day. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001.

day), the increase was 0.0019 standard deviations, or 0.99% (p < 0.01; column (3)). The effect of receiving platform-initiated nudges on the nudge recipient's production was not significant on Day 4 (the third day after the first reception day; column (4)).

Next we compared the production-boosting effects of social nudges and platform-initiated nudges. Figure 4 displays the relative effect sizes of these two kinds of nudges, as well as the corresponding 95% confidence intervals.[6] The effect of receiving platform-initiated nudges on *Number of Videos Uploaded* was generally below that of receiving social nudges. In particular, on the reception day (i.e., $t = 1$), receiving social nudges increased the number of uploaded videos by 13.21% (p < 0.0001), more than twice as large as the increase of 5.55% (p < 0.0001) engendered by receiving platform-initiated nudges; on the day following the first reception day (i.e., $t = 2$), receiving social nudges increased the number of uploaded videos by 5.29% (p < 0.0001), almost three times as large as the increase of 1.37% (p < 0.0001) brought by receiving platform-initiated nudges; on the second day following the first reception day (i.e., $t = 3$), receiving social nudges increased the number of uploaded videos by 2.54% (p < 0.01), almost twice as large as the increase of 0.99% (p < 0.01) from by receiving platform-initiated nudges.[7]

[6] The upper (lower) bound of each 95% confidence interval in Figure 4 equaled the upper (lower) bound of the 95% confidence interval of the corresponding regression coefficient on treatment (based on raw data) divided by the average of the outcome variable in the control condition.

[7] Notably, during the four days since the first reception day (including the first reception day itself) for which we reported the day-by-day effects of both nudges, most (88%) providers were sent only one social nudge in the social

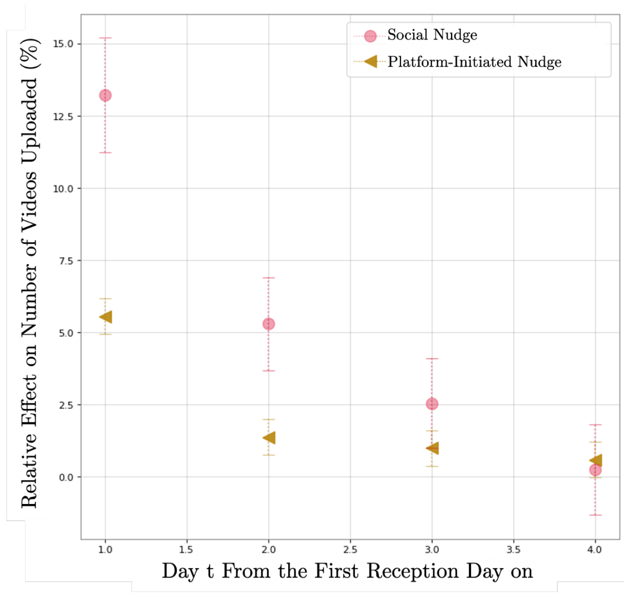


Figure 4 Comparing the Relative Effect Size of Receiving a Platform-Initiated Nudge Versus Social Nudge Over Time

Note: The error bars represent 95% confidence intervals.

In addition, as a robustness check, we analyzed only providers who were included in both the platform-nudge experiment and our first social-nudge experiment ($N = 63,467$) to as cleanly estimate the difference between these two kinds of nudges as possible. Among these overlapping providers, we re-analyzed the direct effects of receiving social nudges or platform-initiated nudges. Since the effects of social nudges and platform-initiated nudges on the recipients' production among these overlapping providers were no longer statistically significant after the first reception day, we focused on comparing these effects on the first reception day. As shown in Table 18 Panel B, the effect of receiving platform-initiated nudges on the number of videos uploaded on the first reception day was not significant ($p = 0.10$; column (1))[8], while receiving social nudges significantly boosted the number of videos uploaded on the first reception day by 0.0216 standard deviations, or 12.35% ($p < 0.001$; column (2)). These results further provide suggestive evidence that social nudges boosted providers' production to a larger extent than platform-initiated nudges.

Nevertheless, platform-initiated nudges like what Platform O tested can still be quite useful to platforms, as it is not an easy job to develop interventions that improve online platforms' operational performance. For instance, at Google and Bing, more than 10,000 online field experiments

nudge experiment, and most (89%) providers were sent only one platform-initiated nudge in the platform-initiated nudge experiment. Hence, the production-boosting effects of both social nudges and platform-initiated nudges were mostly driven by one nudge, making the comparison fair.

[8] Even if we put aside whether the estimated effect was statistically significant, receiving platform-initiated nudges was estimated to increase the number of videos uploaded on the first reception day by 0.0152 standard deviations (or 4.54%) among the overlapping sample of providers, which was still lower than the effect of receiving social nudges on the number of videos uploaded on the first reception day (i.e., 0.0216 standard deviations or 12.35%).

are conducted each year, only 10%–20% of which identify interventions with positive effects.[9] As shown above, though less effective than social nudges in boosting production, platform-initiated nudges tested on Platform O did increase production, and the magnitude of their production-boosting effect is actually comparable with the effect sizes of other subtle interventions that recent research tested on online platforms via randomized field experiments. For example, presenting users with different kinds of performance feedback on a Chinese mobile-app-based recipe crowdsourcing platform increased recipe postings by 1.8%-4.8% (Huang et al. 2019), and displaying the number of people who had applied for a job on LinkedIn increased the job application rate by 3.5% (Gee 2019).

D. Proofs for the Social Network Model

This section presents the proofs of our technical results for the social network model, as well as the details for the estimation strategy discussed in Section 6.

D.1. On Condition $\mathcal{C}$

Lemma 2 *If* $(\alpha_d, \mathbf{D})$ *satisfies Condition* $\mathcal{C}$*, the matrix* $\mathbf{I} - (1/(1-\alpha_d))\mathbf{D})$ *is then invertible with*

$$\left(\mathbf{I} - \frac{1}{1-\alpha_d} \cdot \mathbf{D}\right)^{-1} = \lim_{k\to+\infty} \mathbf{M}(k) = \lim_{k\to+\infty}\left(\mathbf{I} + \sum_{i=1}^{k} \frac{1}{(1-\alpha_d)^i}\mathbf{D}^i\right). \tag{14}$$

Proof of Lemma 2. By condition $\mathcal{C}$, $\lim\limits_{k\to+\infty} \mathbf{M}(k)$ exists, which we denote as $\mathbf{M}$. Therefore,

$$\lim_{k\to+\infty} \frac{1}{(1-\alpha_d)^k} \cdot \mathbf{D}^k = \mathbf{0}_{|E|\times|E|}, \text{ where } \mathbf{0}_{|E|\times|E|} \text{ is the 0 matrix of dimension } |E| \times |E|.$$

Furthermore, we have that

$$\left(\mathbf{I} - \frac{1}{1-\alpha_d} \cdot \mathbf{D}\right)\mathbf{M} = \lim_{k\to+\infty}\left(\mathbf{I} - \frac{1}{1-\alpha_d} \cdot \mathbf{D}\right)\mathbf{M}(k) = \lim_{k\to+\infty}\left(\mathbf{I} - \frac{1}{(1-\alpha_d)^{k+1}} \cdot \mathbf{D}^{k+1}\right) = \mathbf{I},$$

where the second inequality follows from the identity $(\mathbf{I}-\mathbf{A})(\mathbf{I}+\mathbf{A}+\mathbf{A}^2+\cdots+\mathbf{A}^k) = \mathbf{I}-\mathbf{A}^{k+1}$ for any square matrix $\mathbf{A}$. Therefore, $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D}$ is invertible and its inverse is given by Equation (14). □

Remark 1 *The proof of Lemma 2 also implies that, under Condition* $\mathcal{C}$*,* $\mathbf{I}-\mathbf{D}$*, is invertible with*

$$(\mathbf{I}-\mathbf{D})^{-1} = \mathbf{I} + \sum_{k=1}^{+\infty} \mathbf{D}^k \geq \mathbf{0}_{|E|\times|E|}. \tag{15}$$

[9] See https://hbr.org/2017/09/the-surprising-power-of-online-experiments for details.

Condition $\mathcal{C}$ is sometimes hard to verify in practice, so we leverage the concept of matrix norm to provide an easy-to-verify sufficient condition. Specifically, we denote ℓ_q-norm of matrices by $||\cdot||_q$ for any $q \in [1, +\infty]$, which is the operator norm defined through $||\mathbf{A}||_q = \sup_{\boldsymbol{z}:||\boldsymbol{z}||_q \le 1} ||\mathbf{A}\boldsymbol{z}||_q$ for any squared matrix $\mathbf{A}$ and $\boldsymbol{z}$ with appropriate dimensions (Horn and Johnson 2012). Also, we say that the $(\alpha_d, \mathbf{D})$ satisfies $\mathcal{C}_q(\delta)$ for some $\delta \in (0,1)$, provided that $||(1/(1-\alpha_d))\mathbf{D}||_q \le \delta$. We note that if $(\alpha_d, \mathbf{D})$ satisfies $\mathcal{C}_q(\delta)$ for some $\delta \in (0,1)$, the inverse of $\mathbf{I} - (1/(1-\alpha_d))\mathbf{D}$ is given by Equation (14) (see, e.g., Corollary 5.6.16 and Corollary 5.6.17 of Horn and Johnson 2012), which also implies that Condition $\mathcal{C}$ is satisfied.

For Platform O, based on our estimates of $\mathbf{D}$ and α_d (see Section 6.2, Table 8 in particular), we quantify that $(\alpha_d, \mathbf{D})$ satisfies $\mathcal{C}_\infty(0.6)$.[10] Therefore, Condition $\mathcal{C}$ is satisfied for Platform O.

D.2. Proof of Theorem 1

Let us assume throughout the proof that $(\alpha_d, \mathbf{D})$ satisfies Condition $\mathcal{C}$. Also recall that the system of the social network model is defined by Equations (3) and (4).

Let us denote by $\boldsymbol{y}^*(t) := \mathbb{E}[\boldsymbol{y}(t)]$ and $\boldsymbol{x}^*(t) := \mathbb{E}[\boldsymbol{x}(t)]$. Since $\epsilon_e^y(t)$ and $\epsilon_e^x(t)$ are the random errors with a zero mean, it then follows from Equation (4) that $\boldsymbol{y}^*(t) = \boldsymbol{\mu} + \sum_{1\le s\le t-1} \alpha_d^{t-s}\mathbf{D}\boldsymbol{y}^*(s) + \mathbf{D}\boldsymbol{y}^*(t)$, or equivalently

$$\boldsymbol{y}^*(t) = (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1\le s\le t-1} \alpha_d^{t-s}\mathbf{D}\boldsymbol{y}^*(s)\right), \tag{16}$$

where $(\mathbf{I}-\mathbf{D})^{-1}$ is well-defined by Equation (15). We also note that $\boldsymbol{y}^*(1) = (\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu}$. Similarly, by Equation (3), we have:

$$x_i^*(t) = \sum_{1\le s\le t-1} \alpha_p^{t-s} \sum_{e\in E: e_d = i} p_e y_e^*(s). \tag{17}$$

We now show that the sequence $\{\boldsymbol{y}^*(t) : t \ge 1\}$ is componentwise increasing and bounded, so it converges to a limit. We relegate the proof of the following lemma after the proof of Theorem 1.

Lemma 3 *Regarding $\{\boldsymbol{y}^*(t) : t \ge 1\}$, the following statements hold:*

(a) *For each $e \in E$, $y_e^*(t)$ is increasing in t ($t \in \mathbb{Z}_+$).*

(b) *For each $e \in E$ and each $t \in \mathbb{Z}_+$, $y_e^*(t) \le \mathcal{BE}_e(\mathbf{D}, \boldsymbol{\mu})$.*

Lemma 3 implies that the limit of the sequence $\{\boldsymbol{y}^*(t) : t \ge 1\}$ exists and is finite, which we denote as $\boldsymbol{y}^*$. Note that, by Equation (16),

$$\boldsymbol{y}^*(t+1) = (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1\le s\le t} \alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) \tag{18}$$

[10] To protect the sensitive data, we cannot report the tightest value of δ.

and

$$\alpha_d \boldsymbol{y}^*(t) = (\mathbf{I}-\mathbf{D})^{-1}\left(\alpha_d\boldsymbol{\mu} + \sum_{1\le s\le t-1}\alpha_d^{t-s+1}\mathbf{D}\boldsymbol{y}^*(s)\right). \tag{19}$$

Taking the difference between Equation (18) from Equation (19) leads to $\boldsymbol{y}^*(t+1) - \alpha_d\boldsymbol{y}^*(t) = (\mathbf{I}-\mathbf{D})^{-1}\left((1-\alpha_d)\boldsymbol{\mu} + \alpha_d\mathbf{D}\boldsymbol{y}^*(t)\right)$. Taking $t\to+\infty$ on both sides leads to $\boldsymbol{y}^* - \alpha_d\boldsymbol{y}^* = (\mathbf{I}-\mathbf{D})^{-1}\left((1-\alpha_d)\boldsymbol{\mu}+\alpha_d\mathbf{D}\boldsymbol{y}^*\right)$. Reorganizing the terms, we have $\boldsymbol{y}^* = \boldsymbol{\mu} + (1/(1-\alpha_d))\mathbf{D}\boldsymbol{y}^*$. Since $\mathbf{I}-(1/(1-\alpha_d))\mathbf{D}$ is invertible by Lemma 2, $\boldsymbol{y}^* = (\mathbf{I}-(1/(1-\alpha_d))\mathbf{D})^{-1}\boldsymbol{\mu} = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$.

Finally, it remains to show that $\lim_{t\to\infty}\mathbb{E}[x(t)] = x^*$ and $x^* = \boldsymbol{\eta}^T\boldsymbol{y}^*$. With the same argument as the proof for $\boldsymbol{y}^*$, we have, by Equation (17), $\lim_{t\to\infty}\mathbb{E}[\boldsymbol{x}(t)] = \boldsymbol{x}^*$ where $x_i^* = \frac{1}{1-\alpha_p}\sum_{e\in E: e_d = i} p_e y_e^*$, for all $i\in V$. Therefore,

$$x^* = \sum_{i\in V} x_i^* = \sum_{e\in E}\frac{1}{1-\alpha_p}p_e y_e^* = \boldsymbol{\eta}^T\boldsymbol{y}^*.$$

To avoid repetition, we omit the details and conclude the proof of Theorem 1. □

We now give the proof of Lemma 3.

Proof of Lemma 3. We prove **Part (a)** by induction. Note that $\boldsymbol{y}^*(2) = (\mathbf{I}-\mathbf{D})^{-1}(\boldsymbol{\mu}+\alpha_d\mathbf{D}\boldsymbol{y}^*(1))$. Because $y_e^*(1)\ge 0$ for all $e\in E$, we have, by Equation (15),

$$\boldsymbol{y}^*(2) = (\mathbf{I}-\mathbf{D})^{-1}(\boldsymbol{\mu}+\alpha_d\mathbf{D}\boldsymbol{y}^*(1)) \ge (\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu} = \boldsymbol{y}^*(1).$$

Therefore, the base case holds.

Next, we show that if $y_e^*(s)\ge y_e^*(s-1)$ for all $e\in E$ and $2\le s\le t$, then $y_e^*(t+1)\ge y_e^*(t)$ for all $e\in E$. By Equation (16), we have

$$\begin{aligned}
\boldsymbol{y}^*(t+1)-\boldsymbol{y}^*(t) =& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu}+\sum_{1\le s\le t}\alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) - (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu}+\sum_{1\le s\le t-1}\alpha_d^{t-s}\mathbf{D}\boldsymbol{y}^*(s)\right)\\
=& (\mathbf{I}-\mathbf{D})^{-1}\left(\alpha_d^t\mathbf{D}\boldsymbol{y}^*(1) + \sum_{s=1}^{t-1}\alpha_d^s\mathbf{D}(\boldsymbol{y}^*(t+1-s)-\boldsymbol{y}^*(t-s))\right)\\
\ge& \mathbf{0}_{|E|},
\end{aligned}$$

where the inequality follows from the inductive hypothesis, $y_e^*(1)\ge 0$ for all $e\in E$, and $(\mathbf{I}-\mathbf{D})^{-1}\ge \mathbf{0}_{|E|\times|E|}$ (see Equation (15)). Therefore, **Part (a)** follows immediately from the standard induction argument.

We prove **Part (b)** by induction. Observe that, by Equations (14)-(15) and $0<\alpha_d<1$,

$$\boldsymbol{y}^*(1) = (\mathbf{I}-\mathbf{D})^{-1}\boldsymbol{\mu} = \left(\mathbf{I}+\sum_{k=1}^{+\infty}\boldsymbol{D}^k\right)\boldsymbol{\mu} \le \left(\mathbf{I}+\sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\boldsymbol{D}^k\right)\boldsymbol{\mu} = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu}).$$

The base case holds.

Next, we show that if $\boldsymbol{y}^*(s) \leq \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ for all $1 \leq s \leq t$, it holds that $\boldsymbol{y}^*(t+1) \leq \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$. By Equations (15)-(16) and the inductive hypothesis, we have

$$\begin{aligned}
\boldsymbol{y}^*(t+1) =& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1\leq s\leq t} \alpha_d^{t+1-s}\mathbf{D}\boldsymbol{y}^*(s)\right) \\
\leq& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \sum_{1\leq s\leq t} \alpha_d^{t+1-s}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right) \\
=& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d(1-\alpha_d^t)}{1-\alpha_d}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right) \\
\leq& (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d}{1-\alpha_d}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right),
\end{aligned} \tag{20}$$

where the second inequality follows from $0 < \alpha_d^t < 1$. Expanding $\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ by Equation (14), we have

$$\begin{aligned}
(\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d}{1-\alpha_d}\mathbf{D}\cdot\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})\right) &= (\mathbf{I}-\mathbf{D})^{-1}\left(\boldsymbol{\mu} + \frac{\alpha_d}{1-\alpha_d}\mathbf{D}\cdot\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}\right) \\
&= (\mathbf{I}-\mathbf{D})^{-1}\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{\alpha_d}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}.
\end{aligned} \tag{21}$$

Furthermore, invoking Equation (14), we evaluate $(\mathbf{I}-\mathbf{D})\,\mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$ as follows:

$$\begin{aligned}
(\mathbf{I}-\mathbf{D})\,\mathcal{BE}(\mathbf{D},\boldsymbol{\mu}) &= (\mathbf{I}-\mathbf{D})\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu} \\
&= \left(\mathbf{I} + \sum_{k=1}^{+\infty}\left(\frac{1}{(1-\alpha_d)^k} - \frac{1}{(1-\alpha_d)^{k-1}}\right)\mathbf{D}^k\right)\boldsymbol{\mu} \\
&= \left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{\alpha_d}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}.
\end{aligned} \tag{22}$$

Therefore,

$$\mathcal{BE}(\mathbf{D},\boldsymbol{\mu}) = (\mathbf{I}-\mathbf{D})^{-1}\left(\mathbf{I} + \sum_{k=1}^{+\infty}\frac{\alpha_d}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu}. \tag{23}$$

Combining Equations (20), (21), and (23) immediately yields our desired inequality that $\boldsymbol{y}^*(t+1) \leq \mathcal{BE}(\mathbf{D},\boldsymbol{\mu})$. This proves the induction step. We conclude the proof of Lemma 3. □

Proof of Corollary 1. Since $(\alpha_d, \mathbf{D})$ satisfies Condition $\mathcal{C}$,

$$\lim_{k\uparrow+\infty}\tilde{\boldsymbol{y}}(k) = \lim_{k\uparrow+\infty}\widetilde{\mathcal{BE}}(\mathbf{D},\boldsymbol{\mu},k) = \lim_{k\uparrow+\infty}\mathbf{M}(k)\boldsymbol{\mu} = \mathcal{BE}(\mathbf{D},\boldsymbol{\mu}) = \boldsymbol{y}^*,$$

where the third equality follows from 2 and the fourth from Theorem 1. Hence, we have

$$\lim_{k\uparrow+\infty}\tilde{x}(k) = \lim_{k\uparrow+\infty}\boldsymbol{\eta}^T\tilde{\boldsymbol{y}}(k) = \boldsymbol{\eta}^T\boldsymbol{y}^* = x^*.$$

Since $\mathbf{D} \geq \mathbf{0}$, $\mathbf{M}(k)$ is componentwise increasing in k, so is $\tilde{\boldsymbol{y}}(k) = \mathbf{M}(k)\boldsymbol{\mu}$. Therefore, $\tilde{x}(k) = \boldsymbol{\eta}^T\tilde{\boldsymbol{y}}(k)$ increasing in k as well. □

Algorithm 1 Approximate Global Effect of Social Nudges

Down Sampling: Uniformly randomly sample a subset of nodes $\tilde{V} \subset V$. Find the set of edges that point to a node in $\tilde{V}$, $\tilde{E} := \{e \in E : e_d \in \tilde{V}\}$), and the set of edges that originate from a node in $\tilde{V}$, $\tilde{L} := \{\ell \in E : \ell_o \in \tilde{V}\}$).

Parameter Initialization: For each $e \in \tilde{E}$, estimate μ_e and p_e. For each $\ell \in \tilde{L}$, estimate p_ℓ. For each $e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o$, estimate $d_{e\ell}$. Estimate α_d and α_p.

Direct Effect of Social Nudges on Content Production: Estimate

$$\hat{w}_0 := \sum_{i \in \tilde{V}} \sum_{e \in \tilde{E}, e_d = i} \frac{\mu_e p_e}{1 - \alpha_p}.$$

Indirect Effect of Social Nudges on Content Production: Estimate:

$$\hat{w}_1 := \sum_{i \in \tilde{V}} \sum_{e \in \tilde{E}, \ell \in \tilde{L}, e_d = \ell_o = i} \frac{\mu_e d_{e\ell} p_\ell}{(1 - \alpha_p)(1 - \alpha_d)}$$

Total Production Boost on the Entire Population: Scaling the estimates back to V:

$$\hat{w} := \frac{|V|}{|\tilde{V}|}(\hat{w}_0 + \hat{w}_1)$$

D.3. Estimating the Global Effect of Social Nudges

In this section, we present an approximation algorithm to estimate the global effect of social nudges on production boost. Then, we show that this algorithm generates consistent estimate for $\tilde{x}(1)$. Based on the two approximations (confining the diffusion radius to 1 and downsampling the user nodes to $\tilde{V}$) introduced in Section 6.2, Algorithm 1 provides a detailed global effect estimation procedure. We are now ready to prove that Algorithm 1 produces an unbiased estimate for $\tilde{x}(1)$.

Proposition 1 *Algorithm 1 yields an unbiased estimate for $\tilde{x}(1)$.*

Proof of Proposition 1. As shown in Section 6.2, the total production boost can be theoretically approximated as $\tilde{x}(1)$ by including all providers on the entire social network (the node set V), and is practically approximated as $\hat{w}$ according to our Algorithm 1 by sampling a subset of providers from V (i.e., $\tilde{V}$). Now we prove that $\hat{w}$ is an unbiased estimate for $\tilde{x}(1)$.

First, reorganizing the sum by nodes rather than edges, we have

$$\begin{aligned}\tilde{x}(1) &= \boldsymbol{\eta}^\top \left(\mathbf{I} + \frac{1}{(1-\alpha_d)}\mathbf{D}\right)\boldsymbol{\mu} = \sum_{e \in E}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in E:\ \ell_d = e_o} \frac{\mu_e d_{\ell e} p_e}{(1-\alpha_p)(1-\alpha_d)}\right) \\ &= \sum_{e \in E}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in E:\ e_d = \ell_o} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right) = \sum_{i \in V}\left(\sum_{e \in E:\ e_d = i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in E:\ e_d = \ell_o} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right).\end{aligned}$$

Note that $\hat{w}$ is defined as

$$\hat{w} = \frac{|V|}{|\tilde{V}|}\sum_{i \in \tilde{V}}\left(\sum_{e \in \tilde{E}:\ e_d = i}\left(\frac{\mu_e p_e}{1-\alpha_p} + \sum_{\ell \in \tilde{L}:\ e_d = \ell_o} \frac{\mu_e d_{e\ell} p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right).$$

where $\tilde{E} := \{e \in E : e_d \in \tilde{V}\}$ and $\tilde{L} := \{\ell \in E : \ell_o \in \tilde{V}\}$. Clearly, we can replace $\tilde{E}$ and $\tilde{L}$ with E respectively in the definition of w. Moreover, for each node i in V, we use a binary random variable $s_i \in \{0,1\}$ to denote whether node i is selected in the sample $\tilde{V}$. Then, we can write w as

$$\begin{aligned}\hat{w} &= \frac{|V|}{|\tilde{V}|}\sum_{i\in\tilde{V}}\left(\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p}+\sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell}p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right)\\ &= \frac{|V|}{|\tilde{V}|}\sum_{i\in V}\left(s_i\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p}+\sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell}p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right).\end{aligned}$$

Since we uniformly and randomly sample the node set $\tilde{V}$, we have, for each $i\in V$, $\mathbb{E}[s_i] = \frac{|\tilde{V}|}{|V|}$. Hence,

$$\begin{aligned}\mathbb{E}[\hat{w}] &= \frac{|V|}{|\tilde{V}|}\sum_{i\in V}\left(\mathbb{E}[s_i]\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p}+\sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell}p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right)\\ &= \frac{|V|}{|\tilde{V}|}\sum_{i\in V}\left(\frac{|\tilde{V}|}{|V|}\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p}+\sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell}p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right)\\ &= \sum_{i\in V}\left(\sum_{e\in E:\ e_d=i}\left(\frac{\mu_e p_e}{1-\alpha_p}+\sum_{\ell\in E:\ e_d=\ell_o}\frac{\mu_e d_{e\ell}p_\ell}{(1-\alpha_p)(1-\alpha_d)}\right)\right) = \tilde{x}(1).\end{aligned}$$

This concludes the proof. □

D.4. Production Boost from Nudges Sent by New Users

In this section, we quantity the total production boost attributed to the organic nudges sent by new users. See Online Appendix F for notations and modeling details.

Proposition 2 *The additional production boost per period from the organic nudges sent by new users in N is given by:*

$$\Delta x^* := \bar{x}^* - x^* = \bar{\boldsymbol{\eta}}^T\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{\mu}}_{E'}) = \sum_{e\in E'}\nu_e, \tag{24}$$

provided that $(\alpha_d,\bar{\mathbf{D}})$ satisfies Condition $\mathcal{C}$. Here $\bar{x}^*$ and x^* are the vectors of production boost in the long-run steady state for the network with and without the set of new users, respectively.

Proof of Proposition 2. Define $\bar{\boldsymbol{\mu}}_E \in \mathbb{R}^{|\bar{E}|}$ with $\bar{\mu}_e = \mu_e$ for $e\in E$ and $\bar{\mu}_e = 0$ for $e\in E'$ (recall that $\bar{E} = E\cup E'$ and $E\cap E' = \emptyset$). Hence, $\bar{\boldsymbol{\mu}} = \bar{\boldsymbol{\mu}}_E + \bar{\boldsymbol{\mu}}_{E'}$. By Equation (5), $\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{v}})$ is linear in vector $\bar{\boldsymbol{v}}\in\mathbb{R}^{|\bar{E}|}$. Therefore,

$$\bar{x}^* = \bar{\boldsymbol{\eta}}^T\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{\mu}}) = \bar{\boldsymbol{\eta}}^T\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{\mu}}_E+\bar{\boldsymbol{\mu}}_{E'}) = \bar{\boldsymbol{\eta}}^T\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{\mu}}_E)+\bar{\boldsymbol{\eta}}^T\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{\mu}}_{E'}).$$

Hence, to prove Equation (24), it suffices to show that

$$x^* = \bar{\boldsymbol{\eta}}^T\mathcal{BE}(\bar{\mathbf{D}},\bar{\boldsymbol{\mu}}_E),$$

or, equivalently by Theorem 1,

$$\bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E) = \boldsymbol{\eta}^T \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}). \tag{25}$$

Since N is the set of new users, no one has followed them yet. Therefore, there is no edge whose destination is the origin of any edge in E', i.e., for any $e \in E'$ and any $\ell \in \bar{E}$, we have $\ell_d \neq e_o$. Therefore, by definition, $d_{\ell e} = 0$ for any $e \in E'$ and any $\ell \in \bar{E}$. This implies that $(\bar{\mathbf{D}}\bar{\boldsymbol{\mu}}_E)_e = (\mathbf{D}\boldsymbol{\mu})_e$ for all $e \in E$ and $(\bar{\mathbf{D}}\bar{\boldsymbol{\mu}}_E)_e = 0$ for $e \in E'$. Therefore, by a standard induction argument, we have, for any $k \geq 1$,

$$\left(\frac{1}{(1-\alpha_d)^k}\bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E\right)_e = \left(\frac{1}{(1-\alpha_d)^k}\mathbf{D}^k \boldsymbol{\mu}\right)_e \text{ if } e \in E; \; \left(\frac{1}{(1-\alpha_d)^k}\bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E\right)_e = 0 \text{ if } e \in E'. \tag{26}$$

Because $\bar{\boldsymbol{\eta}}_e = \boldsymbol{\eta}_e$ for all $e \in E$, Equation (26) further implies that, for all $k \geq 1$,

$$\frac{1}{(1-\alpha_d)^k}\bar{\boldsymbol{\eta}}^T \bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E = \frac{1}{(1-\alpha_d)^k}\boldsymbol{\eta}^T \mathbf{D}^k \boldsymbol{\mu}. \tag{27}$$

Furthermore, because $\bar{\mu}_e = 0$ for $e \in E'$, we have

$$\bar{\boldsymbol{\eta}}^T \cdot \mathbf{I} \cdot \bar{\boldsymbol{\mu}}_E = \bar{\boldsymbol{\eta}}^T \bar{\boldsymbol{\mu}}_E = \boldsymbol{\eta}^T \boldsymbol{\mu} = \boldsymbol{\eta}^T \cdot \mathbf{I} \cdot \boldsymbol{\mu}. \tag{28}$$

Therefore, Definition 1 implies that:

$$\begin{aligned}\bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_E) =& \bar{\boldsymbol{\eta}}^T \left(\mathbf{I} + \sum_{k=1}^{+\infty} \frac{1}{(1-\alpha_d)^k}\bar{\mathbf{D}}^k\right)\bar{\boldsymbol{\mu}}_E = \bar{\boldsymbol{\eta}}^T \cdot \mathbf{I} \cdot \bar{\boldsymbol{\mu}}_E + \sum_{k=1}^{+\infty} \frac{1}{(1-\alpha_d)^k}\bar{\boldsymbol{\eta}}^T \bar{\mathbf{D}}^k \bar{\boldsymbol{\mu}}_E \\ =& \boldsymbol{\eta}^T \cdot \mathbf{I} \cdot \boldsymbol{\mu} + \sum_{k=1}^{+\infty} \frac{1}{(1-\alpha_d)^k}\boldsymbol{\eta}^T \bar{\mathbf{D}}^k \boldsymbol{\mu} = \boldsymbol{\eta}^T \left(\mathbf{I} + \sum_{k=1}^{+\infty} \frac{1}{(1-\alpha_d)^k}\mathbf{D}^k\right)\boldsymbol{\mu} = \boldsymbol{\eta}^T \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}),\end{aligned} \tag{29}$$

where the third equality follows from Equations (27) to (28). Therefore, equality (25) holds. Finally, the last equality of (24) follows immediately from the linearity of the BCE measure $\mathcal{BE}(\bar{\mathbf{D}}, \cdot)$ in $\boldsymbol{v}$ and the definition of ν_e (Equation (9)). We have concluded the proof of Proposition 2. □

E. Social Network Model Estimation Details

In this section, we present the estimation details for the global effect of social nudges, including the parameters needed in Algorithm 1: μ_e ($e \in E$), p_e ($e \in E$), $d_{e\ell}$ ($e, \ell \in E$ and $e_d = \ell_o$), α_p, and α_d. We also present two robustness checks.

E.1. Estimation of μ_e

Due to Platform O's rule that a user can send no more than one social nudge to another user each day, estimated μ_e (and $d_{e\ell}$) falls between 0 and 1. In this case, the parameter μ_e measures the expected probability of e_o sending a social nudge to e_d per day when e_o has not received nudges from her followers recently. We estimate μ_e by taking advantage of the fact that providers in the control group of our social nudge experiment cannot receive nudges (and thus cannot be motivated to send more nudges out because of receiving nudges themselves) during the experiment. We sampled 5 million edges uniformly at random from all edges whose origin was in the control condition of our social-nudge experiment. Here, we do not require the origins of these edges to satisfy the selection criteria of our analysis sample mentioned in Section 3 since we use this random edge sample to represent the overall edges on Platform O. Our goal is to train a prediction model to estimate μ_e for each $e \in E$.

We fit the logistic regression model (30) to predict μ_e, i.e., the probability that $\textit{Social-Nudge Incidence}_e = 1$. We select features based on the commonly recognized characteristics in the network economics literature (see, e.g., Jackson 2010) such as the degrees of a node in V (measured by the number of followers and the number of followings the node has) and the strength of an edge in E (measured by whether e_o and e_d has a bi-directional relationship, i.e., whether there exists $e' \in E$ such that $e'_o = e_d$ and $e'_d = e_o$). Among a large set of network-based features that we explore, our final logistic regression model includes features that satisfy two criteria: (1) the coefficient on the feature is statistically significant, and (2) the combination of selected features maximizes the performance of the logistic regression model. Specifically, the final retained features include (1) whether e_o's number of followers was greater than the median value across all origin nodes in the sample (*Large Number of Followers for* e_o), (2) whether e_o's number of followings was greater than the median value across all origin nodes in the sample (*Large Number of Following for* e_o), (3) whether e_d was also following e_o ($\textit{Two-Way Tie}_e$), and (4) the baseline productivity (which equals the average number of videos uploaded per day across the 30 days before the experiment) of e_d (*Baseline Productivity of* e_d)[11].

$$\begin{aligned}&\log\left(\frac{\mathbb{P}(\textit{Social-Nudge Incidence}_e = 1)}{1-\mathbb{P}(\textit{Social-Nudge Incidence}_e = 1)}\right)\\ =&\beta_0 + \beta_1 \textit{Large Number of Followers for } e_o + \beta_2 \textit{Large Number of Following for } e_o\\ &+\beta_3 \textit{Two-Way Tie}_e + \beta_4 \textit{Baseline Productivity of } e_d + \epsilon_e\end{aligned} \tag{30}$$

Table 19 reports the estimated coefficients (β_i) and the standard errors of the estimates. We implement a five-fold cross validation to evaluate the performance of this logistic regression model, which has a 99.99% average accuracy and a 0.78 Area Under Curve (AUC), suggesting qualified prediction performance. For all the edges in $\tilde{E}$, we can estimate the probability that e_o will nudge e_d in a given period by Equation (31).

$$\begin{aligned}\frac{1}{\mu_e} =& 1+\exp(-(\beta_0 + \beta_1 \textit{Large Number of Followers for } e_o + \beta_2 \textit{Large Number of Following for } e_o\\ &+\beta_3 \textit{Two-Way Tie}_e + \beta_4 \textit{Baseline Productivity of } e_d))\end{aligned} \tag{31}$$

[11] The correlation between the baseline productivity and social-nudge incidence is -0.0021.

Table 19 The Results of a Logistic Regression Model Predicting Social-Nudge Incidence

	Coefficient (1)	Standard Error (2)
Intercept	-9.9943****	0.1204
Large Number of Followers for e_o	1.4398****	0.1309
Large Number of Following for e_o	-0.8518****	0.1013
Two-Way Tie$_e$	1.0309****	0.1048
Baseline Productivity of e_d	-0.3977****	0.0951

Note: *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

E.2. Estimation of p_e and α_p

Recall that the parameter p_e ($e \in E$) measures the immediate positive effect of receiving *one* social nudge from e_o on provider e_d's production, i.e., the production boosting effect in the same time period when the nudge is sent. The time discounting factor α_p indicates that receiving one social nudge from e_o boosts provider e_d's production by $p_e\alpha_p^t$ in the t^{th} period after e_d receives the nudge. To cleanly estimate p_e and α_p, from the analysis sample of our social-nudge experiment (as defined in Section 3), we identify $962,120$ providers who were sent only *one* social nudge on their first reception day (accounting for 97% of the analysis sample). Those providers were not sent any nudges prior to the experiment (per the selection criteria of our analysis sample), so they were sent one social nudge for the first time on their first reception day.

Since we jointly estimate parameters p_e and α_p, we focus on estimating p_e as the average treatment effect. That is, p_e is independent of the edge $e \in E$. Specifically, we first estimate the coefficient on treatment (i.e, β_1) in regression specification (1) for each day t since the first reception day (where $t = 1$ refers to the first reception day itself) until β_1 becomes statistically insignificant on a given day t. The dependent variable examined here is *Number of Videos Uploaded*$_{it}$. On Day 4 (i.e., three days after the first reception day), β_1 is no longer statistically significant, so we use the estimates of β_1 from Day 1 to Day 3. The regression results are reported in Table 20.

We denote $p(t)$ as the regression coefficient on treatment estimated using raw data without standardization for Day t ($t = 1, 2, 3$). In Table 20, we report the corresponding regression coefficient on treatment using standardized data to protect Platform O's sensitive information. We jointly estimate (p_e, α_p) by minimizing the sum of squared errors in the following nonconvex program:

$$\min_{(p_e,\alpha_p)} \left\{ \sum_{t=1}^{3} \epsilon_t^2 \;\middle|\; p(t) = p_e\alpha_p^{t-1} + \epsilon_t, t = 1, 2, 3 \right\}. \tag{32}$$

Solving (32) yields p_e and α_p, which we report in Table 8 column (1).

Table 20 Over-Time Direct Effects of Receiving One Social Nudge on Content Production

Panel A: Main Experiment					
Outcome Variable	Number of Videos Uploaded				
	on Day 1 (First Reception Day)	on Day 2	on Day 3	on Day 4	
	(1)	(2)	(3)	(4)	
Treatment	0.0263 ****	0.0125****	0.0077****	0.0003	
	(0.0020)	(0.0020)	(0.0020)	(0.0020)	
Relative Effect Size	13.71%	5.28%	3.12%		
Observations	962,120	962,120	962,120	962,120	
Panel B: Replication Experiment					
Outcome Variable	Number of Videos Uploaded				
	on Day 1 (First Reception Day)	on Day 2	on Day 3	on Day 4	on Day 5
	(1)	(2)	(3)	(4)	(5)
Treatment	0.0237****	0.0183****	0.0102****	0.0060*	0.0028
	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
Relative Effect Size	12.68%	8.56%	5.01%	2.98%	
Observations	655,001	655,001	655,001	655,001	655,001

Note: Number of Videos Uploaded was standardized to have a unit deviation before entering the regressions. Panel A includes providers who were sent only one social nudge on their first reception day in the main experiment. Panel B includes providers who were sent only one social nudge on their first reception day in the replication experiment. The unit of analysis for all columns was a provider on Day t, where $t = 1$ refers to the first reception day. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

E.3. Estimation of $d_{e\ell}$ and α_d

The estimation of $d_{e\ell}$ and α_d follows a similar approach to that of (p_e, α_p). The parameter $d_{e\ell}$ measures the increase in e_d's probability of sending a social nudge to ℓ_d on the day of receiving *one* social nudge from e_o ($e_d = \ell_o$). By definition, $d_{e\ell} = 0$ if $e_d \neq \ell_o$. The parameter α_d quantifies the time-discounting factor of such effect, such that receiving one social nudge from e_o boosts the number of nudges provider e_d would send to ℓ_d by $d_{e\ell}\alpha_d^t$ in the t^{th} period after e_d receives the nudge. We focus on the subset of providers from the analysis sample of our social-nudge experiment who (1) were sent only *one* social nudge on their first reception day and (2) were following at least one user the day before the main experiment. We estimate the diffusion effect of a social nudge by comparing the number of nudges providers sent per following relationship between the treatment and control conditions on and after their first reception day, adopting regression specification (1) for each day since the first reception day.

Consistent with the estimation of p_e and α_p, we jointly estimate parameters $d_{e\ell}$ and α_d. Due to the joint estimation, we focus on estimating $d_{e\ell}$ as the average treatment effect. That is, $d_{e\ell}$ is independent of the edge $e, \ell \in E$. Specifically, we first estimate the coefficient on treatment (i.e., β_1) in regression specification (1) for each day t since the first reception day (where $t = 1$ refers to the first reception day) until β_1 becomes statistically insignificant on a given day t. The dependent variable examined here is *Number of Social Nudges Sent per Edge*$_{it}$. It equals the number of social nudges sent by provider $i \in V$ on day t since the first reception day divided by her number of following. Starting from Day 3, β_1 is no longer statistically significant, so we use the estimates of β_1 from Day 1 to Day 2. The regression results are reported in Table 21.

We denote $d(t)$ as the regression coefficient estimated using raw data without standardization for Day t $(t = 1, 2)$. In Table 21, we report the corresponding regression coefficient on treatment using standardized data to protect Platform O's sensitive information. We jointly estimate $(d_{e\ell}, \alpha_d)$ by minimizing the sum of squared errors in the following nonconvex program:

$$\min_{(d_e, \alpha_d)} \left\{ \sum_{t=1}^{2} \epsilon_t^2 \;\middle|\; d(t) = d_{e\ell}\alpha_d^{t-1} + \epsilon_t, t = 1, 2 \right\}. \tag{33}$$

Solving (33) yields $d_{e\ell}$ and α_d, which we report in Table 8 column (1).

Table 21 Over-Time Diffusion Effects of Receiving One Social Nudge

Panel A: Main Experiment			
Outcome Variable	Number of Social Nudges Sent per Edge		
	on Day 1 (First Reception Day)	on Day 2	on Day 3
	(1)	(2)	(3)
Treatment	0.0080***	0.0049*	0.0025
	(0.0021)	(0.0021)	(0.0021)
Relative Effect Size	10.55%	8.06%	
Observations	947,730	947,730	947,730

Panel B: Replication Experiment			
Outcome Variable	Number of Social Nudges Sent per Edge		
	on Day 1 (First Reception Day)	on Day 2	on Day 3
	(1)	(2)	(3)
Treatment	0.0082***	0.0050*	0.0010
	(0.0025)	(0.0025)	(0.0025)
Relative Effect Size	10.89%	9.06%	
Observations	640,920	640,920	640,920

Notes: Number of Social Nudges Sent per Edge was standardized to have a unit standard deviation before entering the regressions. Panel A includes providers who were sent only one social nudge on their first reception day in the main experiment and were following at least one user the day before the main experiment. Panel B includes providers who were sent only one social nudge on their first reception day in the replication experiment and were following at least one user the day before the replication experiment. The unit of analysis for all columns was a provider on Day t, where $t = 1$ refers to the first reception day. Robust standard errors are reported in the parentheses. *p<0.05; **p<0.01; ***p<0.001; ****p<0.0001

E.4. Robustness Checks

As the first robustness check, we re-sample $\tilde{V}$ and re-estimate the global effect of social nudges using parameters estimated from our main social-nudge experiment. As shown in Table 22, the estimation results based on the new sample of $\tilde{V}$ are very similar to the results based on the sample reported in Section 6.2, confirming that our estimates reported in Table 9 are robust.

As the second robustness check, we re-estimate (p_e, α_p) and $(d_{e\ell}, \alpha_d)$ based on the same method described above but use data from the second social-nudge replication experiment. Regarding the

Table 22 Estimation of the Global Effect of Social Nudges

	The Estimation Result			
	Reported in Section 6.2 (1)		Based on Another Sample of $\tilde{V}$ (2)	
Direct Effect	130.08	*One Day:* 47.55 *Beyond One Day:* 82.53	132.85	*One Day:* 48.56 *Beyond One Day:* 84.29
Indirect Effect	10.59		10.87	
Global Effect	140.67		143.72	
The Ratio of Indirect Effect to Direct Effect	8.14%		8.19%	

parameters p_e and α_p, we estimate the coefficient on treatment (i.e., β_1) in regression specification (1) for each day t where outcome variable is *Number of Videos Uploaded*$_{it}$ since the first reception day (where $t = 1$ refers to the first reception day itself) until β_1 becomes statistically insignificant on a given day t. For the replication experiment, on day 5, β_1 is no longer statistically significant. So we use the estimates of β_1 from Day 1 to Day 4 to jointly estimate p_e and α_p. The regression results using standardized data are presented in Panel B of Table 20. The corresponding solution to the nonconvex program (32) yields p_e and α_p, which we report in Table 8 column (2) and are consistent with the estimates derived from the first experiment.

Regarding the parameters $d_{e\ell}$ and α_d, we estimate the coefficient on treatment (i.e, β_1) in regression specification (1) for each day t where outcome variable is *Number of Social Nudges Sent per Edge*$_{it}$ since the first reception day (where $t = 1$ refers to the first reception day itself) until β_1 becomes statistically insignificant on a given day t. Starting from Day 3, β_1 is no longer statistically significant, so we use the estimates of β_1 from Day 1 to Day 2 to jointly estimate $d_{e\ell}$ and α_d. The regression results are presented in Panel B of Table 21. The corresponding solution to the nonconvex program (33) yields $d_{e\ell}$ and α_d, which we report in Table 8 column (2) and are consistent with the estimates derived from the first experiment.

In addition, we apply Algorithm 1 and the data from the second social-nudge experiment to estimate the global effect of social nudges. We report the estimation results in Table 9 column (2). Compared to the naïve estimation, including the over-time accumulation of the direct boosting effect of social nudges on recipients' production leads to a 200% (i.e., (146.06 − 48.65)/48.65) increase, and considering nudge diffusion leads to an additional 25% (i.e., 12.24/48.65) increase. The indirect production boost from social-nudge diffusion accounts for at least 8.38% (i.e., 12.24/146.06) of the direct production effect. All of these results are fairly consistent with our estimation results based on data from the first social-nudge experiment (see Table 9).

F. Operational Problems About Social Network Model

In this section, we study two operational problems with our social network model: (1) optimal seeding, and (2) provider recommendation for new users. We solve the problems leveraging the SNI developed in Section 6.3.

F.1. Optimal Seeding

Here we provide details about how to solve the optimal seeding application as presented in Section 6.3. Assuming that user e_o will, on average, send more nudges to e_d if the platform encourages her to do so, we denote that for each $e \in K$, the average number of social nudges sent per day will increase by a relative effect of δ_μ after e_o receives the motivation from the platform (i.e., from μ_e to $\mu_e(1+\delta_\mu)$).[12]

It is straightforward to derive that the global effect increment of social nudges with respect to the selected edges, K, is $\boldsymbol{\eta}^T \mathcal{BE}(\mathbf{D}, \boldsymbol{\mu}_K)\delta_\mu$, where $\boldsymbol{\mu}_K \in \mathbb{R}^{|E|}$ represents a vector with an entry of edge $e \in K$ (resp. $e \notin K$) equal to μ_e (resp. 0). Such production boost can be reasonably approximated by

$$\Delta(K,1) := \boldsymbol{\eta}^T \cdot \widetilde{\mathcal{BE}}(\mathbf{D}, \boldsymbol{\mu}_K, 1)\delta_\mu = \delta_\mu \cdot \sum_{e \in K} \tilde{\nu}_e(1). \tag{34}$$

Therefore, it is (approximately) "optimal" to select n edges in E with the highest (approximate) SNIs, i.e., the n edges with the largest $\tilde{\nu}_e(1)$'s.

Table 23 Global Effect of Social Nudges for Different Strategies

	Setting	Δ_o (1)	Δ_r (2)	Ξ (3)
(1)	$\lvert K\rvert = 0.1\lvert E\rvert$, $\delta_\mu = 100\%$	54.77	13.78	297%
(2)	$\lvert K\rvert = 0.1\lvert E\rvert$, $\delta_\mu = 10\%$	5.48	1.38	297%

The platform may adopt the random strategy, which randomly targets a subset of edges $K \subset E$ and encourages the users to nudge more on these edges. This random strategy is the most straightforward and simplest way to stimulate social nudges sent on a platform. We benchmark the approximately "optimal" strategy (i.e., to target the edges with the highest $\tilde{\nu}_e(1)$'s) against the random strategy and compare their performances in production boost. We define Δ_r (resp. Δ_o) as the additional production boost under the random (resp. "optimal") strategy, and $\Xi := (\Delta_o - \Delta_r)/\Delta_r \times 100\%$ as the relative improvement of the "optimal" strategy over the random strategy. We evaluate these two strategies based on the same sample of $\tilde{V}$ as the one used to generate the

[12] Our method of deriving the optimization strategy can be easily carried over to a setting where the average number of social nudges sent per day will increase by an absolute effect of δ_μ after e_o receives one push from the platform (i.e., from μ_e to $\mu_e + \delta_\mu$) for each $e \in K$.

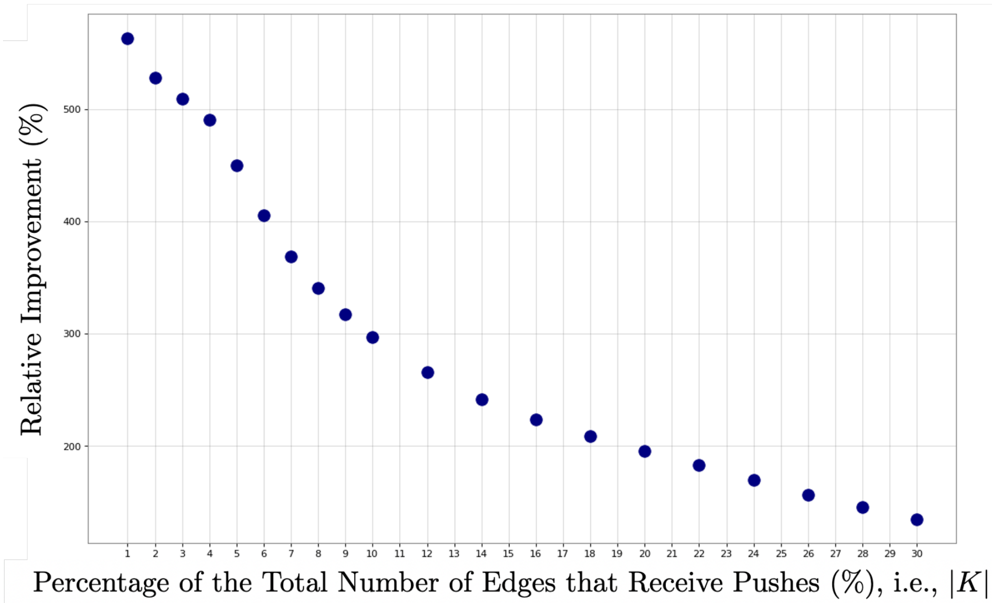


Figure 5 Relative Improvements of the Optimal Strategy Over the Random Strategy as $|K|$ Changes

global effect estimates in Table 9 column (1). We also examine how Ξ changes according to δ_μ and the size of the targeted edges $n = |K|$. The simulation results are reported in Table 23 and Figure 5. The primary observation is that the "optimal" strategy based on (approximate) SNIs substantially outperforms the random strategy, regardless of the effectiveness of the platform's encouragement for users to send additional nudges (i.e., the magnitude of δ_μ). In particular, this relative edge is most prominent if the constraint on the number of targeted edges is tighter (i.e., n is smaller).

F.2. Content Provider Recommendation for New Users

Below, we provide details about the provider recommendation problem for newly registered platform users. The platform will first construct a provider list based on the basic demographic features (e.g., age, gender and location) of each new user, together with her interests in specific content categories reported upon registration. Next, the platform decides the ranking of the provider list, following which it sequentially recommends the content providers to the new user.

Recall that the existing social network of Platform O is denoted by $G = (V, E)$, where V is the set of existing nodes (users), and E is the set of existing edges (following relationships). We denote the set of newly registered users as N. For each new user $i \in N$, denote the set of existing providers this user chooses to follow as U_i and the associated set of new following relationships as $E_i := \{(i,u) : u \in U_i\}$. Therefore, the new social network with those new users can be written as $\bar{G} := \left(\bar{V}, \bar{E}\right)$, where $\bar{V} := V \cup N$ is the set of all users and $\bar{E} := E \cup \left(\bigcup_{i \in N} E_i\right)$ is the set of all edges. The production boost vector $\bar{\boldsymbol{\eta}} := (p_e/(1-\alpha_p) : e \in \bar{E})$, the organic nudge vector $\bar{\boldsymbol{\mu}} := (\mu_e : e \in \bar{E})$, and the diffusion matrix $\bar{\mathbf{D}} := (d_{\ell e} : (\ell, e) \in \bar{E}^2)$ with respect to the new social network can be defined accordingly. By Theorem 1, as long as $\bar{\mathbf{D}}$ satisfies Condition $\mathcal{C}$, the total number of nudges and the total production boost per period are given by $\bar{\boldsymbol{y}}^* = \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}})$ and $\bar{x}^* := \bar{\boldsymbol{\eta}}^T \bar{\boldsymbol{y}}^*$, respectively. Define $E' := \bigcup_{i \in N} E_i$ as the set of new edges and $\bar{\boldsymbol{\mu}}_{E'} \in \mathbb{R}^{|\bar{E}|}$ as a vector with an entry of edge $e \in E'$ (resp. $e \notin E'$) equal to μ_e (resp. 0). The additional production boost attributed to the social nudges sent by the new users is given by $\Delta x^* = \bar{\boldsymbol{\eta}}^T \mathcal{BE}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_{E'}) = \sum_{e \in E'} \nu_e$.[13] Such additional production boost can be reasonably approximated by

$$\Delta \tilde{x}^*(1) := \bar{\boldsymbol{\eta}}^T \cdot \widetilde{\mathcal{BE}}(\bar{\mathbf{D}}, \bar{\boldsymbol{\mu}}_{E'}, 1) = \sum_{e \in E'} \tilde{\nu}_e(1) = \sum_{i \in N} \left(\sum_{e \in E_i} \tilde{\nu}_e(1) \right), \tag{35}$$

provided that $\tilde{\nu}_e(1)$ is a reasonable approximation of ν_e for each $e \in E'$. Note that Equation (35) also implies the content provider recommendation of each new user can be optimized separately.

Let us now consider the content provider recommendation problem for the new user $i \in N$. To begin with, Platform O identifies a content provider list for the new user based on her features, which we denote as $M_i \subset V$. Then, the platform selects $V_i \subset M_i$ with $|V_i| = m$ and recommends the providers in V_i to the user in a sequential manner. Similar to the optimal seeding problem, the platform avoids overly interfering its users, so the total number of recommended providers to each new user, m, is generally not too large but at the magnitudes of a few dozens. Denote the probability that a new user will follow the j-th provider recommended to her as c_j. For simplicity, we assume that c_j is only dependent on the ranking of the provider in the list, i.e. j, but independent of the identity of him. Because a new user will have a higher chance to follow the provider recommended to her earlier, it holds that $c_1 \geq c_2 \geq \cdots \geq c_m$. The platform's recommendation strategy for user i can be summarized as the provider list V_i together with a bijection $\pi : \{1, 2..., m\} \to V_i$, where $\pi(j)$ refers to the provider ranked in the j-th position. By Equation (35), under the recommendation strategy (V_i, π), the (approximate) additional production boost from the social nudges sent by new user i is given by

$$\sum_{j=1}^{m} c_j \tilde{\nu}_{(i,\pi(j))}(1), \tag{36}$$

where $(i, \pi(j))$ is the edge in E_i, representing that new user i follows the j-th recommended provider and, thus, $c_j \tilde{\nu}_{(i,\pi(j))}(1)$ is the expected additional production boost by recommending the j-th provider. For any $i' \in M_i$, we call $\tilde{\nu}_{(i,i')}(1)$ the induced (approximate) SNI of provider i'. It is clear from Equation (36) that the (approximate) "optimal" strategy is to select m providers in M_i with the highest induced (approximate) SNIs and the rank them in the descending order of the induced (approximate) SNI.

Similar to optimal seeding, we would compare the SNI-based provider recommendation with the benchmark random recommendation, which recommends the content providers based on a random

[13] See Proposition 2 in Online Appendix D.4 for the formal result.

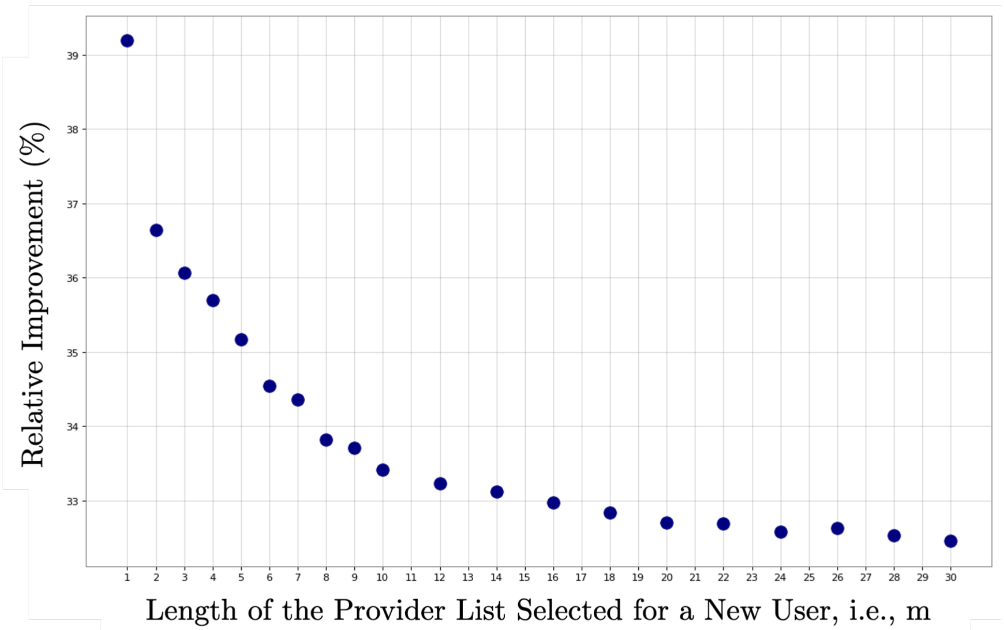


Figure 6 **Relative Improvements of the Optimal Strategy Over the Random Strategy as m Changes**

permutation of M_i. To evaluate the edge of our "optimal" strategy over the random strategy, we randomly sample $1,000$ new users on Platform O and examine the provider list of varying lengths m. We quantify the performance metrics Δ_o (production boost of the "optimal" strategy), Δ_r (production boost of the random strategy), and $\Xi = (\Delta_o - \Delta_r)/\Delta_r \times 100\%$ (the relative improvement of the "optimal" strategy over the random strategy) for recommending providers to new users and report the results in Figure 6. Consistent with the optimal seeding problem, the "optimal" provider recommendation strategy based on SNIs outperforms the random strategy, especially when the provider list length m is small.

In sum, we show that our social network model could help the platform optimize, among others, its seeding and provider recommendation strategies. The platform's (approximately) optimal strategy prescribed by our social-nudge indices is much more effective in boosting total production than the simple random strategy.

G. Data

In this section, we report in several tables the distributional information of the features and outcome variables studied in this paper, correlations between variables, and the distributional information of the degrees of the sample network used in Section 6. We provide a guideline of the organization of these tables, which is intended to help readers locate the information of interest. Specifically, in Tables 25–36, for each variable mentioned in the paper (including the main body and appendices), we first standardize it to have a unit deviation if it is a continuous variable (as explained in Section 3), and then report the quantiles (i.e., 1%, 25%, 50%, 75%, and 99%) in the full sample, as well as mean and standard deviation in each condition (i.e., treatment or control). In Table 37, for variables mentioned in the main body of the paper, we report moderate to large correlations between variables (i,e, cases where the absolute value of the correlation coefficient is no less than 0.30). In Table 38, we report the quantiles of the in-degrees and out-degrees of the sample of nodes used to calculate the global effect of social nudges in Section 6 (i.e., $\tilde{V}$). Like other variables, we are not allowed to reveal the distributional information of the raw data on the in-degrees or out-degrees. In order to maintain the relative magnitude between the in-degrees and out-degrees of these nodes, we first scale the in-degrees and out-degrees by a fixed constant (rather than their respective standard deviation, which differs between in-degrees and out-degrees), and then calculate the quantiles.

Table Index	Focus	Statistics
Table 25, Table 26	The first social-nudge experiment	Quantiles
Table 27	The second social-nudge experiment	Quantiles
Table 28	The platform-initiated nudge experiment	Quantiles
Table 29	A matched viewer sample for examining cannibalization	Quantiles
Table 30	Providers who were sent only one social nudge	Quantiles
Table 25, Table 31	The first social-nudge experiment	Mean and standard deviation in each condition
Table 33	The second social-nudge experiment	Mean and standard deviation in each condition
Table 34	The platform-initiated nudge experiment	Mean and standard deviation in each condition
Table 35	A matched viewer sample for examining cannibalization	Mean and standard deviation in each condition
Table 36	Providers who were sent only one social nudge	Mean and standard deviation in each condition
Table 37	All variables studied in the main body	Correlation
Table 38	Sample network	Quantiles

Table 24 Summary of Tables Related to Data Disclosure

Statistics Prior to the Experiment						
Variable	Location	1%	25%	50%	75%	99%
Female (Binary)[1]	Table 1	0.0000	0.0000	1.0000	1.0000	1.0000
Number of Followers	Table 1	0.0000	0.0026	0.0090	0.0289	0.6083
Number of Following	Table 1	0.0000	0.1749	0.4680	1.1108	4.3647
Number of Uploaded Videos	Table 1	0.0000	0.0000	0.0000	0.3528	4.5866
Number of Days with Videos Uploaded	Table 1	0.0000	0.0000	0.0000	0.8214	4.1070
Historical Like Rate	Table 3	0.0000	0.3182	0.6869	1.3090	4.3511
Statistics Related to the First Reception Day						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded	Table 2	0.0000	0.0000	0.0000	0.0000	3.6644
Upload Incidence (Binary)	Table 2	0.0000	0.0000	0.0000	0.0000	1.0000
Number of Videos Uploaded Conditional on Uploading Anything[2]	Table 2	1.8322	1.8322	1.8322	3.6644	12.8253
Two-Way Tie (Binary)	Table 2	0.0000	0.0000	0.0000	1.0000	1.0000
Total Views	Table 3	0.0000	0.0000	0.0000	0.0000	5.2715
Complete View Rate	Table 3	0.0000	0.1274	0.6607	1.4905	4.2946
Like Rate	Table 3	0.0000	0.4039	0.9596	1.6964	4.2857
Comment Rate	Table 3	0.0000	0.0000	0.4852	1.1716	4.3539
Following Rate	Table 3	0.0000	0.0000	0.0000	0.0000	4.0533
Number of Social Nudges Sent	Table 5	0.0000	0.0000	0.0000	0.0000	4.9896
Number of Likes on the First Reception Day	Table 15	0.0000	0.0000	0.0000	0.0000	7.4641
Number of Comments on the First Reception Day	Table 15	0.0000	0.0000	0.0000	0.0000	6.2778
Number of Videos Uploaded Among Low-Productivity Providers[3]	Table 17	0.0000	0.0000	0.0000	0.0000	3.6644
Number of Videos Uploaded Among Medium-Productivity Providers[4]	Table 17	0.0000	0.0000	0.0000	1.8322	7.3287
Number of Videos Uploaded Among High-Productivity Providers[5]	Table 17	0.0000	0.0000	1.8322	3.6644	18.3218

Table 25 Summary Statistics About the First (Main) Social-Nudge Experiment (I)

[1]Note: Hereafter, we label all the binary variables. For each binary variable, we maintained its raw values, rather than standardizing it to have a unit deviation.
[2]Note: By definition, *Number of Videos Uploaded Conditional on Uploading Anything* includes the nonzero values of *Number of Videos Uploaded*. We standardized the *Number of Videos Uploaded* across all providers involved in the first social-nudge experiment, and then extracted its nonzero standardized values to construct *Number of Videos Uploaded Conditional on Uploading Anything*.
[3,4,5]Note: For these three variables, we first standardized Number of Videos Uploaded across all providers involved in the first social-nudge experiment, and then extracted the standardized values for low-productivity, medium-productivity, and high-productivity providers, respectively.

Statistics Beyond the First Reception Day						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 2	Table 4	0.0000	0.0000	0.0000	0.0000	4.2770
Number of Videos Uploaded on Day 3	Table 4	0.0000	0.0000	0.0000	0.0000	3.8533
Number of Videos Uploaded on Day 4	Table 4	0.0000	0.0000	0.0000	0.0000	3.9571
Number of Social Nudges Sent on Day 2	Table 6	0.0000	0.0000	0.0000	0.0000	6.0318
Number of Social Nudges Sent on Day 3	Table 6	0.0000	0.0000	0.0000	0.0000	6.0384
Statistics About All Providers Who Were Sent at Least One Nudge in the First Social-Nudge Experiment (Including Those Who Received Nudges Prior to It)						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on the First Reception Day	Table 10	0.0000	0.0000	0.0000	0.0000	4.9110
Number of Social Nudges Sent on the First Reception Day	Table 10	0.0000	0.0000	0.0000	0.0000	6.3290
Statistics Within 24 Hours Following the First Nudge						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded Within 24 Hours Following the First Nudge	Table 11	0.0000	0.0000	0.0000	0.0000	4.2495
Upload Incidence Within 24 Hours Following the First Nudge (Binary)	Table 11	0.0000	0.0000	0.0000	0.0000	1.0000
Statistics About Control Providers Used to Explore Whether Being Blocked from Social Nudges Caused Reactance						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded	Table 14	0.0000	0.0000	0.0000	0.0000	4.5216
Private Messages Incidence (Binary)	Table 14	0.0000	0.0000	0.0000	0.0000	1.0000
First Reception Day (Binary)	Table 14	0.0000	0.0000	0.5000	1.0000	1.0000
Statistics Broken Down by Whether Providers Received Any Private Message From their First Social-Nudge Sender						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded Among Providers Who Received Private Messages From the First Social-Nudge Sender	Table 14	0.0000	0.0000	0.0000	0.0000	7.3287
Number of Videos Uploaded Among Providers Who Received No Private Messages From the First Social-Nudge Sender	Table 14	0.0000	0.0000	0.0000	0.0000	3.6644

Table 26 Summary Statistics About the First (Main) Social-Nudge Experiment (II)

Statistics During the Experimental Period						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 12	0.0000	0.0000	0.0000	0.0000	3.7439
Number of Videos Uploaded on Day 2	Table 12	0.0000	0.0000	0.0000	0.0000	4.8366
Number of Videos Uploaded on Day 3	Table 12	0.0000	0.0000	0.0000	0.0000	5.0023
Number of Videos Uploaded on Day 4	Table 12	0.0000	0.0000	0.0000	0.0000	4.8774
Number of Social Nudges Sent on Day 1	Table 13	0.0000	0.0000	0.0000	0.0000	5.1553
Number of Social Nudges Sent on Day 2	Table 13	0.0000	0.0000	0.0000	0.0000	3.5356
Number of Social Nudges Sent on Day 3	Table 13	0.0000	0.0000	0.0000	0.0000	4.2500
Number of Social Nudges Sent on Day 4	Table 13	0.0000	0.0000	0.0000	0.0000	4.4280
Number of Social Nudges Sent on Day 5	Table 13	0.0000	0.0000	0.0000	0.0000	4.4868

Table 27 Summary Statistics About the Second (Replication) Social-Nudge Experiment (I)

Statistics During the Second Social-Nudge Experiment						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 18	0.0000	0.0000	0.0000	0.0000	4.6657
Number of Videos Uploaded on Day 2	Table 18	0.0000	0.0000	0.0000	0.0000	4.7433
Number of Videos Uploaded on Day 3	Table 18	0.0000	0.0000	0.0000	0.0000	4.7279
Number of Videos Uploaded on Day 4	Table 18	0.0000	0.0000	0.0000	0.0000	4.7239
Statistics About Providers Who Were Both in the First Social-Nudge and Platform-Initiated Nudge Experiments						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1 of the Platform-Initiated Nudge Experiment	Table 18	0.0000	0.0000	0.0000	0.0000	6.2209
Number of Videos Uploaded on Day 1 of the First Social-Nudge Experiment	Table 18	0.0000	0.0000	0.0000	0.0000	5.4965

Table 28 Summary Statistics About the Platform-Initiated Nudge Experiment (I)

Statistics About Data Used in the Cannibalization Analysis						
Variable	Location	1%	25%	50%	75%	99%
Number of Likes Marked	Table 16	0.0000	0.0095	0.0476	0.2667	5.2484
Number of Comments Left	Table 16	0.0000	0.0000	0.0098	0.0881	3.1047
Incidence of Sending Social Nudges (Binary)	Table 16	0.0000	0.0000	0.5000	1.0000	1.0000
Post (Binary)	Table 16	0.0000	0.0000	0.5000	1.0000	1.0000

Table 29 Summary Statistics About the Cannibalization Analysis (I)

Statistics During the First Social-Nudge Experiment						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 20	0.0000	0.0000	0.0000	0.0000	3.6644
Number of Videos Uploaded on Day 2	Table 20	0.0000	0.0000	0.0000	0.0000	4.2770
Number of Videos Uploaded on Day 3	Table 20	0.0000	0.0000	0.0000	0.0000	3.8533
Number of Videos Uploaded on Day 4	Table 20	0.0000	0.0000	0.0000	0.0000	3.9571
Number of Social Nudges Sent per Edge on Day 1	Table 21	0.0000	0.0000	0.0000	0.0000	1.4113
Number of Social Nudges Sent per Edge on Day 2	Table 21	0.0000	0.0000	0.0000	0.0000	1.2236
Number of Social Nudges Sent per Edge on Day 3	Table 21	0.0000	0.0000	0.0000	0.0000	0.9718
Statistics During the Second Social-Nudge Experiment						
Variable	Location	1%	25%	50%	75%	99%
Number of Videos Uploaded on Day 1	Table 20	0.0000	0.0000	0.0000	0.0000	3.7439
Number of Videos Uploaded on Day 2	Table 20	0.0000	0.0000	0.0000	0.0000	4.8366
Number of Videos Uploaded on Day 3	Table 20	0.0000	0.0000	0.0000	0.0000	5.0023
Number of Videos Uploaded on Day 4	Table 20	0.0000	0.0000	0.0000	0.0000	3.2516
Number of Videos Uploaded on Day 5	Table 20	0.0000	0.0000	0.0000	0.0000	5.0074
Number of Social Nudges Sent per Edge on Day 1	Table 21	0.0000	0.0000	0.0000	0.0000	1.4468
Number of Social Nudges Sent per Edge on Day 2	Table 21	0.0000	0.0000	0.0000	0.0000	0.9906
Number of Social Nudges Sent per Edge on Day 3	Table 21	0.0000	0.0000	0.0000	0.0000	0.5149

Table 30 Summary Statistics About Providers Who Were Sent Only One Social Nudge During a Given Experiment (I)

Statistics Prior to the Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Female (Binary)	492,599	0.5134	0.4998	492,182	0.5138	0.4998
Number of Followers	496,976	0.0622	1.0383	496,700	0.0605	0.9602
Number of Following	496,976	0.8485	1.0008	496,700	0.8480	0.9992
Number of Uploaded Videos	496,976	0.3674	0.9851	496,700	0.3693	1.0147
Number of Days with Videos Uploaded	496,976	0.5057	0.9977	496,700	0.5078	1.0023
Historical Like Rate	430,522	0.9688	1.0003	430,771	0.9689	0.9997
Statistics Related to the First Reception Day						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded	496,976	0.2241	1.0537	496,700	0.1980	0.9430
Upload Incidence (Binary)	496,976	0.0770	0.2666	496,700	0.0677	0.2511
Number of Videos Uploaded Conditional on Uploading Anything[1]	38,281	2.9099	2.5688	33,602	2.9267	2.2715
Two-Way Tie (Binary)	496,976	0.4577	0.4982	496,700	0.4529	0.4978
Total Views	496,976	0.1816	1.0211	496,700	0.1645	0.9784
Complete View Rate	38,154	0.9526	1.0004	33,480	0.9519	0.9995
Like Rate	38,154	1.1530	0.9963	33,480	1.1703	1.0042
Comment Rate	38,154	0.7934	1.0032	33,480	0.8002	0.9964
Following Rate	38,154	0.2372	1.0152	33,480	0.2331	0.9824
Number of Social Nudges Sent	496,976	0.2412	1.0214	496,700	0.2087	0.9779
Number of Likes on the First Reception Day	496,976	0.2595	1.0071	496,700	0.2483	0.9928
Number of Comments on the First Reception Day	496,976	0.2104	1.0106	496,700	0.1996	0.9893
Number of Videos Uploaded Among Low-Productivity Providers[2]	450,919	0.1358	0.7745	450,367	0.1137	0.6519
Number of Videos Uploaded Among Medium-Productivity Providers[3]	41,791	0.8931	1.8235	42,047	0.8354	1.7681
Number of Videos Uploaded Among High-Productivity Providers[4]	4,266	3.0133	4.5146	4,286	2.7987	4.0303
Statistics Beyond the First Reception Day						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 2	496,976	0.2568	1.0042	496,700	0.2439	0.9958
Number of Videos Uploaded on Day 3	496,976	0.2608	1.0077	496,700	0.2543	0.9922
Number of Videos Uploaded on Day 4	496,976	0.2514	0.9989	496,700	0.2508	1.0011
Number of Social Nudges Sent on Day 2	496,976	0.1910	1.0158	496,700	0.1770	0.9838
Number of Social Nudges Sent on Day 3	496,976	0.1813	1.0037	496,700	0.1785	0.9963

Table 31 Summary Statistics About the First (Main) Social-Nudge Experiment (III)

[1,2,3,4]Note: See the notes in Table 25 for how we constructed these variables.

Statistics Among All Providers Who Were Sent at Least One Nudge in the First Social-Nudge Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on the First Reception Day	973,236	0.2551	1.0310	972,882	0.2329	0.9678
Number of Social Nudges Sent on the First Reception Day	973,236	0.2643	1.0246	972,882	0.2320	0.9745
Statistics Within 24 Hours Following the First Nudge						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded Within 24 Hours Following the First Nudge	496,976	0.2685	1.0339	496,700	0.2388	0.9646
Upload Incidence Within 24 Hours Following the First Nudge (Binary)	496,976	0.1121	0.3155	496,700	0.0990	0.2987
Statistics Based on Private-Message Incidence						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded Among Providers Who Received Any Private Messages From the First Social-Nudge Sender	15,073	0.5729	1.6714	13,069	0.5019	1.5337
Number of Videos Uploaded Among Providers Who Received No Private Messages From the First Social-Nudge Sender	481,903	0.2132	1.0265	483,631	0.1898	0.9204

Table 32 Summary Statistics About the First (Main) Social-Nudge Experiment (IV)

Experimental Statistics						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	338,415	0.2152	1.0034	339,675	0.1924	0.9965
Number of Videos Uploaded on Day 2	338,415	0.2385	1.0175	339,675	0.2213	0.9822
Number of Videos Uploaded on Day 3	338,415	0.2190	0.9913	339,675	0.2107	1.0086
Number of Videos Uploaded on Day 4	338,415	0.2009	0.9727	339,675	0.1976	1.0265
Number of Social Nudges Sent on Day 1	338,415	0.2323	1.0264	339,675	0.1998	0.9727
Number of Social Nudges Sent on Day 2	338,415	0.1732	1.0301	339,675	0.1517	0.9689
Number of Social Nudges Sent on Day 3	338,415	0.1533	1.0141	339,675	0.1449	0.9858
Number of Social Nudges Sent on Day 4	338,415	0.1467	1.0127	339,675	0.1410	0.9872
Number of Social Nudges Sent on Day 5	338,415	0.1438	1.0096	339,675	0.1398	0.9903

Table 33 Summary Statistics About the Second (Replication) Social-Nudge Experiment (II)

Statistics During the Experimental Period						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	5,522,864	0.2004	1.0080	5,520,612	0.1899	0.9919
Number of Videos Uploaded on Day 2	5,522,864	0.1945	1.0057	5,520,612	0.1919	0.9943
Number of Videos Uploaded on Day 3	5,522,864	0.1929	1.0022	5,520,612	0.1910	0.9978
Number of Videos Uploaded on Day 4	5,522,864	0.1921	1.0001	5,520,612	0.1910	0.9999
Statistics Among Providers Who Were Both in the First Social-Nudge and Platform-Initiated Nudge Experiments						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1 of the Platform-Initiated Nudges Experiment	31,752	0.3924	1.3493	31,715	0.3753	1.2725
Number of Videos Uploaded on Day 1 of the Social-Nudges Experiment	31,870	0.2601	1.0957	31,597	0.2315	1.0027

Table 34 Summary Statistics About the Platform-Initiated Nudge Experiment (II)

Statistics about the Observational Data Used in the Cannibalization Analysis						
	Post = 0 &					
	Incidence of Sending Social Nudges = 1			*Incidence of Sending Social Nudges = 0*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Likes Marked	353,041	0.5308	1.1520	353,041	0.2715	0.7699
Number of Comments Left	353,041	0.2907	1.2175	353,041	0.1071	0.6511
	Post = 1 &					
	Incidence of Sending Social Nudges = 1			*Incidence of Sending Social Nudges = 0*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Likes Marked	353,041	0.5807	1.1925	353,041	0.2658	0.7578
Number of Comments Left	353,041	0.3222	1.2725	353,041	0.1063	0.6588

Table 35 Summary Statistics About the Cannibalization Analysis (II)

Statistics in the First Social-Nudge Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	481,376	0.2101	1.0192	480,744	0.1848	0.9071
Number of Videos Uploaded on Day 2	481,376	0.2395	0.9615	480,744	0.2275	0.9550
Number of Videos Uploaded on Day 3	481,376	0.2446	0.9731	480,744	0.2372	0.9513
Number of Videos Uploaded on Day 4	481,376	0.2349	0.9558	480,744	0.2347	0.9618
Number of Social Nudges Sent per Edge on Day 1	474,188	0.0811	0.9868	473,542	0.0733	0.9590
Number of Social Nudges Sent per Edge on Day 2	474,188	0.0649	1.0103	473,542	0.0601	0.9448
Number of Social Nudges Sent per Edge on Day 3	474,263	0.0559	0.9710	473,613	0.0536	0.9483
Statistics in the Second Social-Nudge Experiment						
	Treatment Condition			*Control Condition*		
Variable	Observations	Mean	St. Dev.	Observations	Mean	St. Dev.
Number of Videos Uploaded on Day 1	326,817	0.2009	0.9634	328,184	0.1783	0.9430
Number of Videos Uploaded on Day 2	326,817	0.2216	0.9784	328,184	0.2041	0.9353
Number of Videos Uploaded on Day 3	326,817	0.2050	0.9572	328,184	0.1952	0.9617
Number of Videos Uploaded on Day 4	326,817	0.1883	0.9155	328,184	0.1829	0.9082
Number of Videos Uploaded on Day 5	326,817	0.2018	0.9646	328,184	0.1991	0.9654
Number of Social Nudges Sent per Edge on Day 1	319,822	0.0813	0.9882	321,098	0.0733	0.9516
Number of Social Nudges Sent per Edge on Day 2	319,822	0.0574	0.9484	321,098	0.0527	0.9703
Number of Social Nudges Sent per Edge on Day 3	319,925	0.0444	0.9922	321,194	0.0427	0.9646

Table 36 Summary Statistics About Providers Who Were Sent Only One Social Nudge During a Given Experiment (II)

Variables		Correlation
Number of Uploaded Videos in Prior One Week	Number of Days with Videos Uploaded in Prior One Week	0.8145
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 1	0.3617
Number of Uploaded Videos in Prior One Week	Upload Incidence on Day 1	0.3206
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 1 Conditional on Uploading Anything	0.3167
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 2	0.4259
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 3	0.4555
Number of Uploaded Videos in Prior One Week	Number of Videos Uploaded on Day 4	0.4380
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 1	0.3258
Number of Days with Videos Uploaded in Prior One Week	Upload Incidence on Day 1	0.3673
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 2	0.3813
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 3	0.3945
Number of Days with Videos Uploaded in Prior One Week	Number of Videos Uploaded on Day 4	0.3830
Number of Days with Videos Uploaded in Prior One Week	Total Views	0.3190
Number of Videos Uploaded on Day 1	Upload Incidence on Day 1	0.7558
Number of Videos Uploaded on Day 1	Number of Videos Uploaded on Day 1 Conditional on Uploading Anything	1.0000
Number of Videos Uploaded on Day 1	Number of Videos Uploaded on Day 2	0.3529
Number of Videos Uploaded on Day 1	Total Views	0.5805
Upload Incidence on Day 1	Total Views	0.6197
Number of Videos Uploaded on Day 1 Conditional on Uploading Anything	Number of Videos Uploaded on Day 2	0.3377
Number of Videos Uploaded on Day 2	Number of Videos Uploaded on Day 3	0.4188
Number of Videos Uploaded on Day 2	Number of Videos Uploaded on Day 4	0.3015
Number of Videos Uploaded on Day 3	Number of Videos Uploaded on Day 4	0.4245
Like Rate	Comment Rate	0.4769
Like Rate	Historical Like Rate	0.5857
Comment Rate	Historical Like Rate	0.4253

Table 37 Correlation Between Variables When the Absolute Value of Correlation Is Greater Than 0.30

Notes: We only calculate the correlations between variables mentioned in the main text (Tables 1–6). These variables were measured among the full sample of providers in the first (main) social-nudge experiment. Any pair of variables that is not shown in this table has an absolute value of correlation coefficient below 0.30. For the variables that do not have values among some providers (e.g., *Number of Videos Uploaded on Day 1 Conditional on Uploading Anything*), we ignore the providers who do not values. That is, when we calculate the correlation between two variables, we only consider the providers who have values in both variables.

Statistics About the In-degrees and Out-degrees of the Sample Network									
Variable	0.1%	1%	5%	25%	50%	75%	95%	99%	99.9%
In-degrees	0.0556	0.0556	0.0556	0.0556	0.2222	0.6667	8.0000	49.6111	619.9022
Out-degrees	0.0556	0.0556	0.0556	0.1667	0.7222	2.7222	17.6667	46.8333	53.9444

Table 38 Summary Statistics About the Sample Network